

CHIMICA ANALITICA II

CON LABORATORIO

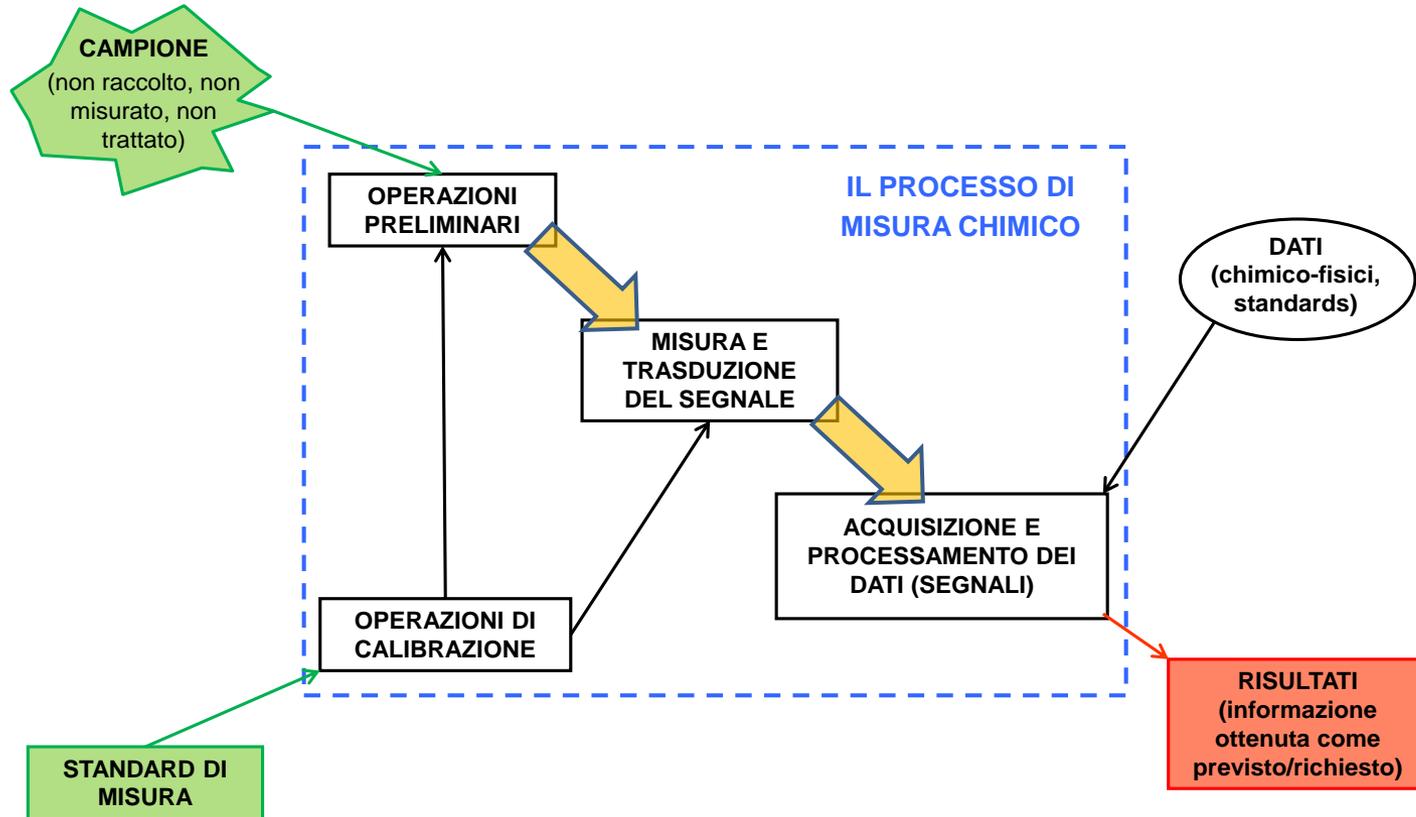
(AA 2021-22)

8 C.F.U. - Laurea triennale in Chimica

Campionamento

INTRODUZIONE

Torniamo allo schema del processo di misura chimico, o processo analitico:



Il CAMPIONAMENTO può essere considerato come il primo stadio delle **OPERAZIONI PRELIMINARI** che, assieme alla **preparazione del campione**, consentono di rendere il campione adatto alla misurazione.

➤ Obiettivi di campionamento

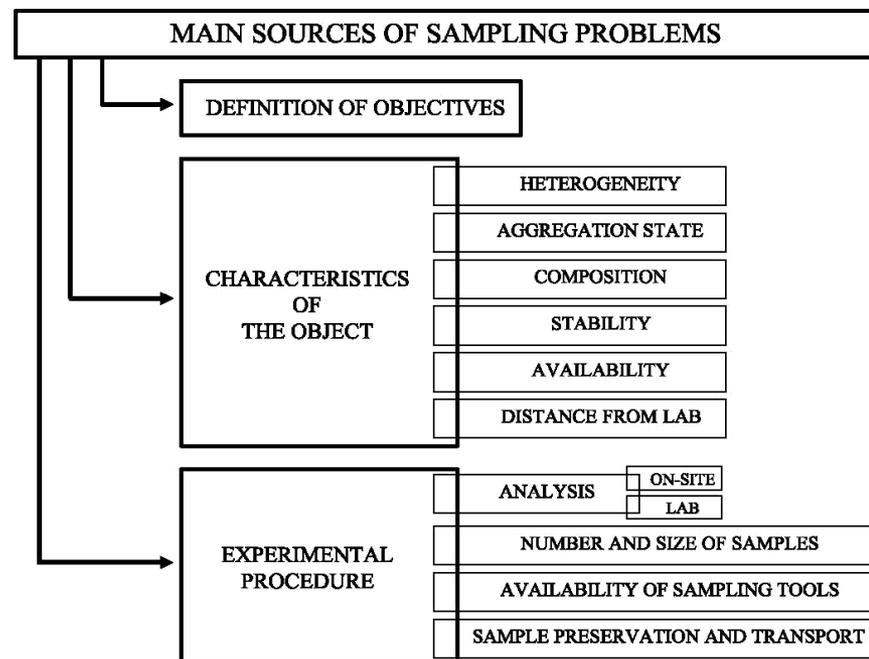
Il campionamento può essere definito in diversi modi, complementari tra di loro:

- la **procedura pratica** utilizzata per selezionare uno a più porzioni (aliquote) del materiale oggetto di studio;
- la linea di collegamento tra l'oggetto e gli altri passaggi della procedura analitica;
- il modo in cui si ottiene la **RAPPRESENTATIVITÀ** del risultato analitico rispetto ad una serie specifica di caratteristiche del materiale oggetto di studio, considerando sia le caratteristiche del materiale che le informazioni che si intende ricavare da esso.

Ottenere un campione rappresentativo del materiale oggetto di indagine non è una operazione semplice, bisogna valutare diversi fattori, ad es.:

- orizzonte spaziale;
- orizzonte temporale;
- stabilità del campione nel tempo;
- quantità di materiale disponibile.

Inoltre le **caratteristiche proprie del materiale e della procedura sperimentale** che si vuole utilizzare per l'analisi condizionano la **procedura di campionamento**.



© 2006 Wiley-VCH, Weinheim
Kellner / Analytical Chemistry
ISBN: 3-527-30590-4 Fig-04-04

Es. Volumi d'aria campionati in ambienti indoor per identificare presenza RNA di SARS-CoV-2 nell'aerosol

Document	Sampling areas	Patients	Sampler	Sampler positioning	Sampling time and flow rate	N of samples	% of positive samples	RNA concentration
Chia et al. 2020	Three airborne infection isolation rooms in the general ward (GW)	5 days from swab tested positive	NIOSH biosampler (different PM fractions)	Two samplers per room (one sample for each bedside – different heights from the ground)	4h at 3.5 L/min	6	66.7 %	from 1.84×10^3 to 3.38×10^3 (copies per m ³)
Guo et al. 2020 ^s	Isolation ward of ICU and GW (patients' rooms, corridor, pharmacy, doctors' office)	ICU (15 patients); GW (24 patients)	SASS 2300 Wetted Wall Cyclone Sampler	Near patients and air outlet	30 min at 300 L/min	65 (approx 10 days campaign)	6 % (intense positive); 18 % (weak positive)	from 0.52 to 3.8 (copies per L)
Jiang et al. 2020 ^s	High-risk area (isolation wards, consulting rooms, observation rooms)	1 positive patient in ICU; 15 suspected positive patients	MAS-100 ECO microbial air sampler	-	100 L/min	28	3.6 %	-
Lei et al. 2020	ICU and GW (patients' rooms)	ICU (4 patients); GW (5 patients)	NIOSH biosampler and DingBlue air sampler	Near the patient's bed; in the patient's bathroom	4h at 3.5 L/min (NIOSH); 30 min at 14 L/min	8 in ICU + 6 in GW (2-4 days repeated samples)	12.5 % (ICU); 33.3 % (GW); 16.6 % (weak-GW)	(Ct from 35.5 to 44.6)

Problema generale (spesso indicato/posto da non specialista)

Definizione degli obiettivi («Scoping») e procedure

Pianificazione raccolta dati/campionamento

(Modello concettuale dell'oggetto / sistema / fenomeno da descrivere)

Campionamento e produzione **dati**/misure

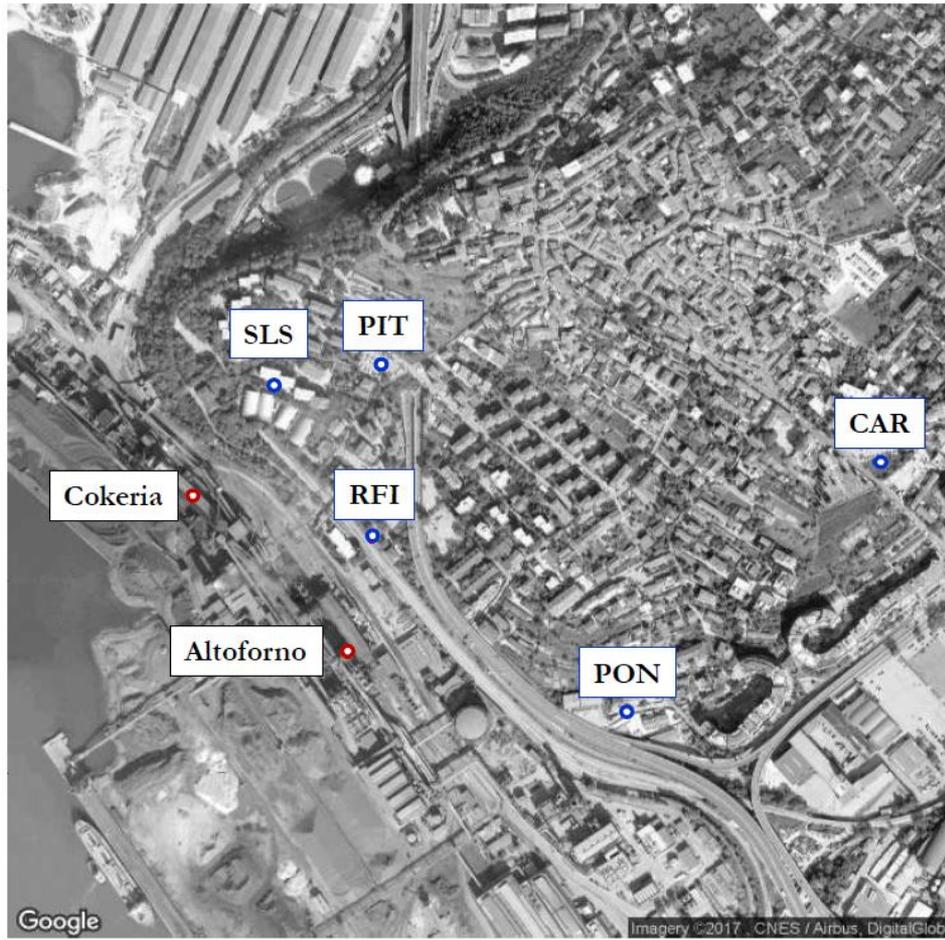
Razionalizzazione/*elaborazione* dati misurati e raccolta dati «ancillari», con produzione di **informazione** (es. concentrazione di X più alta del «normale», qui invece che là, al tempo t_3 invece che al t_7 , associata a concentrazioni basse di Y, che si verificano dopo 40 minuti dall'evento E)

Contestualizzazione dell'informazione chimica prodotta, integrazione di informazioni «non chimico-fisiche» e interpretazione del fenomeno oggetto di studio: produzione di **conoscenza**, utile come *supporto alle decisioni*

Esempio: CARATTERIZZAZIONE DELLA QUALITA' DELL'ARIA

Scoping: Criteri generali minimi (DLgs 155/2010) vs contesto specifico

POSITIONING THE MONITORING STATIONS IN THE RIGHT (Relevant) PLACE



Before 2016:

CAR, the only station under public control, official node of the public AQ monitoring network

RFI, just for industrial performance assessment (not AQ)

PIT, under private control

In 2016:

PON starts to monitor (in Summer 2016)

In 2017:

Continuous noise monitoring close to SLS

SLS starts to monitor, initially for initiative of the Prosecutor's Office of the Court of Trieste

In 2018: SLS under control of ARPA



DLgs 155/2010

Allegato V

(art. 7, commi 1, 2 e 3, e art. 12, comma 2)

Numero minimo delle stazioni di misurazione per biossido di zolfo, biossido di azoto, ossidi di azoto, particolato (PM10 e PM2,5), piombo, benzene, monossido di carbonio, arsenico, cadmio, nichel e benzo(a)pirene.

1. Numero minimo di stazioni di misurazione per la valutazione della qualità dell'aria ambiente in relazione ai valori limite previsti per la protezione della salute umana ed alle soglie di allarme nelle zone e negli agglomerati in cui le misurazioni in siti fissi costituiscono l'unica fonte di informazioni.

1. Per le fonti diffuse si applicano le seguenti tabelle:

Tabella 1

Popolazione dell'agglomerato o della zona (in migliaia di abitanti)	Se la concentrazione massima supera la soglia di valutazione superiore (1) (2)		Se la concentrazione massima è compresa tra la soglia di valutazione superiore e quella inferiore	
	Per inquinanti diversi dal PM	Per il PM (3) (somma delle stazioni di PM10 e PM2,5)	Per inquinanti diversi dal PM	Per il PM (3) (somma delle stazioni di PM10 e PM2,5)
0-249	1	2	1	1
250-499	2	3	1	2
500-749	2	3	1	2
750-999	3	4	1	2
1000-1499	4	6	2	3
1500-1999	5	7	2	3
2000-2749	6	8	3	4
2750-3749	7	10	3	4
3750-4749	8	11	3	6
4750-5999	9	13	4	6
≥6000	10	15	4	7

Allegato IX

(art. 8, commi 3, 4 e 5 e art. 19 commi 8, 12, 13)

Numero minimo di stazioni di misurazione per l'ozono

1. Numero minimo di stazioni di misurazione nelle zone e negli agglomerati in cui le misurazioni in siti fissi costituiscono l'unica fonte di informazioni.

Popolazione (< 1 000)	Agglomerati (stazioni in siti urbani e suburbani) (1)(2)	Altre zone (stazioni siti suburbani e rurali) (1)
< 250		1 ⁽¹⁾
< 500	1 ⁽¹⁾	2 ⁽¹⁾
< 1 000	2 ⁽¹⁾	2 ⁽¹⁾
< 1 500	3	3
< 2 000	3	4
< 2 750	4	5
< 3 750	5	6
> 3 750	1 stazione supplementare per 2 milioni di abitanti	1 stazione supplementare per 2 milioni di abitanti

(1) Deve essere prevista almeno una stazione di misurazione nei siti suburbani, dove può verificarsi la maggiore esposizione della popolazione.

(2) Negli agglomerati per i quali sono previste due o più stazioni di misurazione, almeno il 50% delle stazioni di misurazione deve essere inserito nei siti suburbani.

(3) Nei casi previsti dal paragrafo 4, punto 4 le stazioni di misurazione possono essere assenti alle condizioni ivi previste.

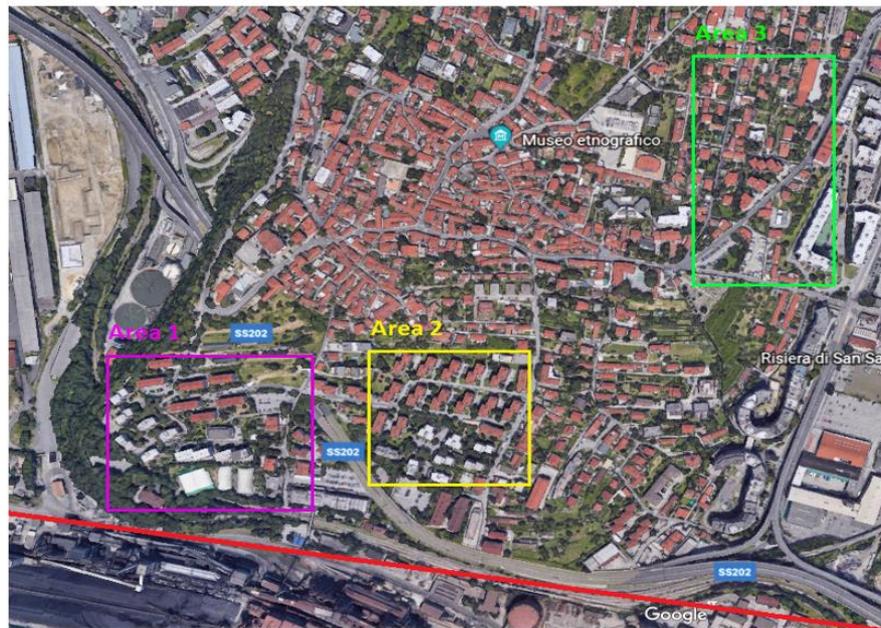


Figura 10 – fotografia satellitare dell'area di Servola vicino alla fiera, con indicazione delle tre aree a cui viene fatto riferimento nel corso dello studio

SERVOLA - ZONA I (via san Lorenzo in Selva, via Pitacco)



INDIRIZZO	NUMERO	Tot-Residenti	Numero-Famiglie	ZONA
PITACCO GIORGIO via 2	14	7	1	1-SELVA
PITACCO GIORGIO via 3	10	5	1	1-SELVA
PITACCO GIORGIO via 4	12	8	1	1-SELVA
PITACCO GIORGIO via 6	14	10	1	1-SELVA
PITACCO GIORGIO via 7	9	8	1	1-SELVA
PITACCO GIORGIO via 8	11	7	1	1-SELVA
PITACCO GIORGIO via 9	11	6	1	1-SELVA
PITACCO GIORGIO via 11	18	9	1	1-SELVA
PITACCO GIORGIO via 11/1	16	8	1	1-SELVA
PITACCO GIORGIO via 11/2	15	7	1	1-SELVA
PITACCO GIORGIO via 11/3	15	8	1	1-SELVA
PITACCO GIORGIO via 11/4	16	11	1	1-SELVA
PITACCO GIORGIO via 11/5	12	9	1	1-SELVA
PITACCO GIORGIO via 11/6	16	8	1	1-SELVA
PITACCO GIORGIO via 13	13	8	1	1-SELVA
PITACCO GIORGIO via 15	12	6	1	1-SELVA
PITACCO GIORGIO via 17	13	7	1	1-SELVA
PITACCO GIORGIO via 19	8	6	1	1-SELVA
PITACCO GIORGIO via 21	10	8	1	1-SELVA
PITACCO GIORGIO via 23	14	7	1	1-SELVA
SAN LORENZO IN SELVA via 21	21	10	1	1-SELVA
SAN LORENZO IN SELVA via 23	5	2	1	1-SELVA
SAN LORENZO IN SELVA via 25	3	2	1	1-SELVA
SAN LORENZO IN SELVA via 25/1	14	8	1	1-SELVA
SAN LORENZO IN SELVA via 25/2	32	14	1	1-SELVA
SAN LORENZO IN SELVA via 25/3	21	12	1	1-SELVA
SAN LORENZO IN SELVA via 25/4	14	8	1	1-SELVA
SAN LORENZO IN SELVA via 27	7	3	1	1-SELVA
SAN LORENZO IN SELVA via 146	15	8	1	1-SELVA
SAN LORENZO IN SELVA via 148	12	11	1	1-SELVA
SAN LORENZO IN SELVA via 150	18	8	1	1-SELVA
SAN LORENZO IN SELVA via 154	8	5	1	1-SELVA



COMUNE DI TRIESTE

Trieste

SERVOLA - ZONA 3 (via Carpineto, via dei Vigneti, Ratto della Pileria)



INDIRIZZO	NUMERO	Tot-Residenti	Numero-Famiglie	ZONA
dei VIGNETI via 46	5	2	1	3-CARPINETO
dei VIGNETI via 48	4	1	1	3-CARPINETO
dei VIGNETI via 50	3	1	1	3-CARPINETO
dei VIGNETI via 56	2	1	1	3-CARPINETO
dei VIGNETI via 58	3	2	1	3-CARPINETO
dei VIGNETI via 58/1	5	1	1	3-CARPINETO
dei VIGNETI via 60	2	1	1	3-CARPINETO
dei VIGNETI via 62	2	2	1	3-CARPINETO
del CARPINETO via 18	31	18	3	3-CARPINETO
del CARPINETO via 20	21	11	3	3-CARPINETO
del CARPINETO via 20/1	29	15	3	3-CARPINETO
del CARPINETO via 21	1	1	1	3-CARPINETO
del CARPINETO via 21/1	2	1	1	3-CARPINETO
del CARPINETO via 22	3	1	1	3-CARPINETO
del CARPINETO via 23	5	3	1	3-CARPINETO
del CARPINETO via 24	2	1	1	3-CARPINETO
del CARPINETO via 24/1	6	4	1	3-CARPINETO
del CARPINETO via 25	1	1	1	3-CARPINETO
del CARPINETO via 26	5	5	1	3-CARPINETO
del CARPINETO via 26/1	2	2	1	3-CARPINETO
del CARPINETO via 26/3	3	1	1	3-CARPINETO
del CARPINETO via 27	9	7	1	3-CARPINETO
del CARPINETO via 27/1	16	7	1	3-CARPINETO
del CARPINETO via 28	5	2	1	3-CARPINETO
del CARPINETO via 29	15	9	1	3-CARPINETO
del CARPINETO via 30	1	1	1	3-CARPINETO
del CARPINETO via 31	6	4	1	3-CARPINETO
del CARPINETO via 33	1	1	1	3-CARPINETO
del CARPINETO via 35	1	1	1	3-CARPINETO
della PILERIA ratto 2	2	1	1	3-CARPINETO
della PILERIA ratto 2/1	4	1	1	3-CARPINETO
della PILERIA ratto 4	5	2	1	3-CARPINETO
della PILERIA ratto 6	4	1	1	3-CARPINETO
della PILERIA ratto 8	3	2	1	3-CARPINETO
della PILERIA ratto 10	4	2	1	3-CARPINETO
della PILERIA ratto 12	3	2	1	3-CARPINETO
VALERIO FESTO via 1	3	3	1	3-CARPINETO
VALERIO FESTO via 3	2	1	1	3-CARPINETO
VALERIO FESTO via 5	2	2	1	3-CARPINETO
VALERIO FESTO via 7	1	1	1	3-CARPINETO
VALERIO FESTO via 9	3	1	1	3-CARPINETO



COMUNE DI TRIESTE

Trieste

(1) **Detecting and characterising sources of persistent organic pollutants (PAHs and PCBs) in surface sediments of an industrialized area (harbour of Trieste, northern Adriatic Sea)**



Gianpiero Adami,^a Pierluigi Barbieri,^a Stefano Piselli,^a Sergio Predonzani^b and Edoardo Reisenhofer^a

^aDepartment of Chemical Sciences, University of Trieste, Via Giorgieri 1, 34127 Trieste, Italy

^bLaboratory of Marine Biology, Via Piccard 54, 34010 Trieste, Italy

Received 7th February 2000, Accepted 30th March 2000

Published on the Web 3rd May 2000

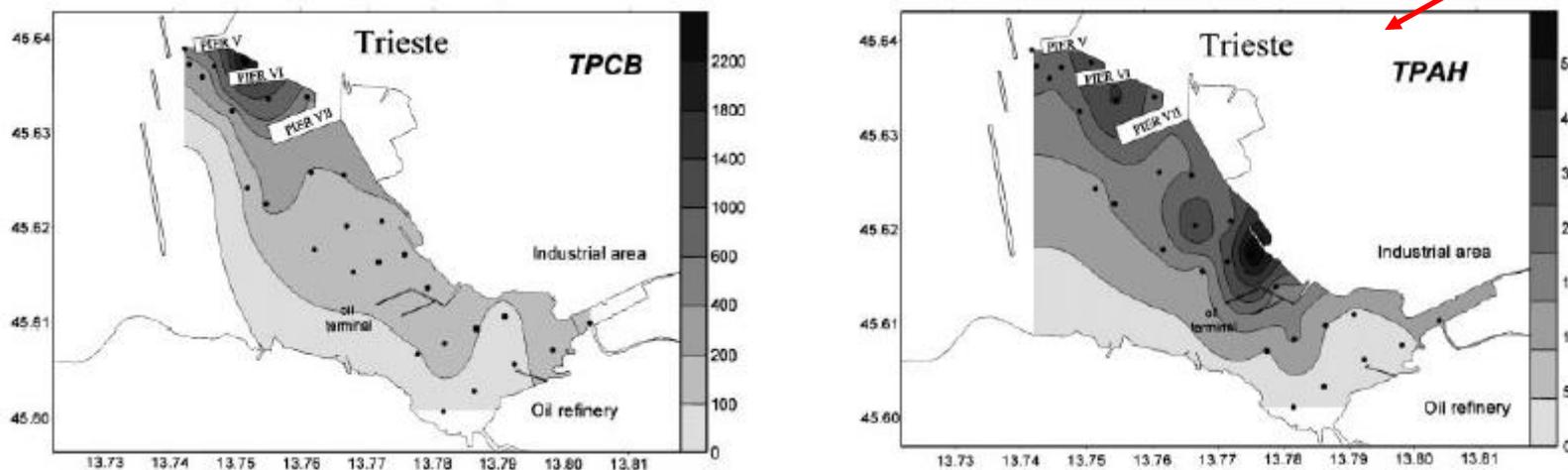
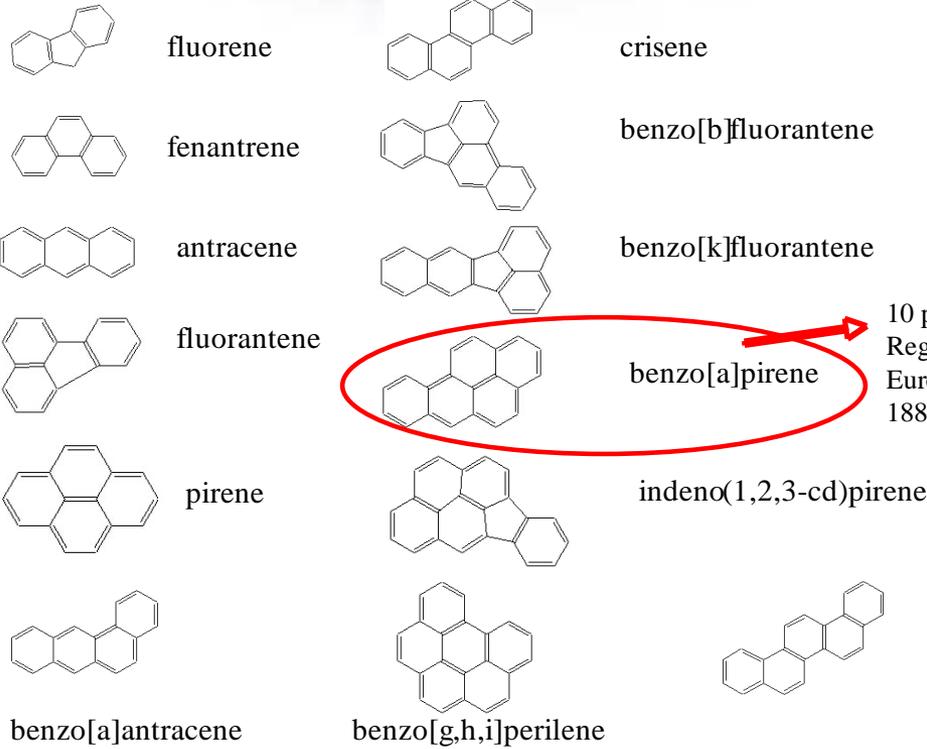
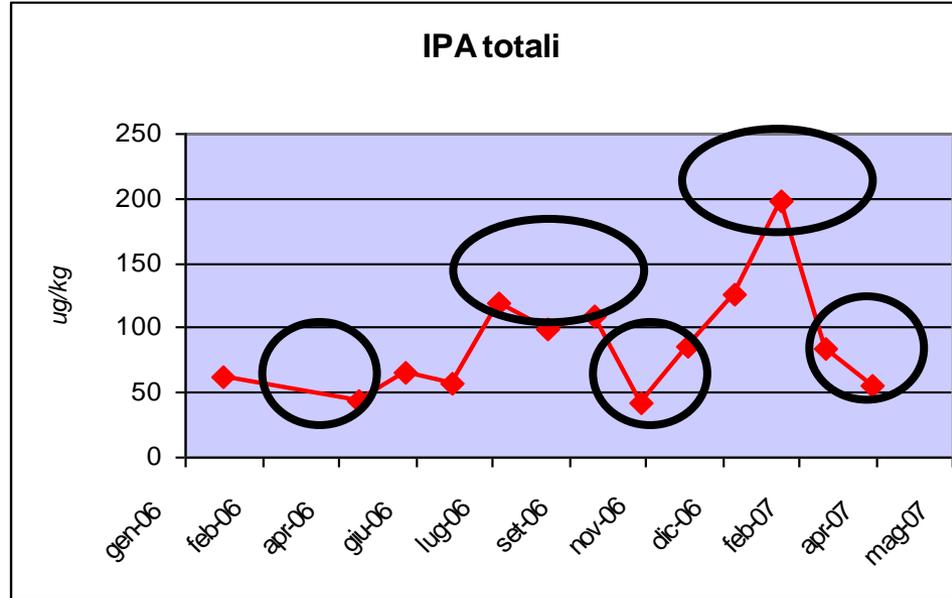


Fig. 4 Iso-concentration curves for (a) TPCB (ng g^{-1} dry sediment) and (b) TPAH ($\mu\text{g g}^{-1}$ dry sediment).

Andamento IPA

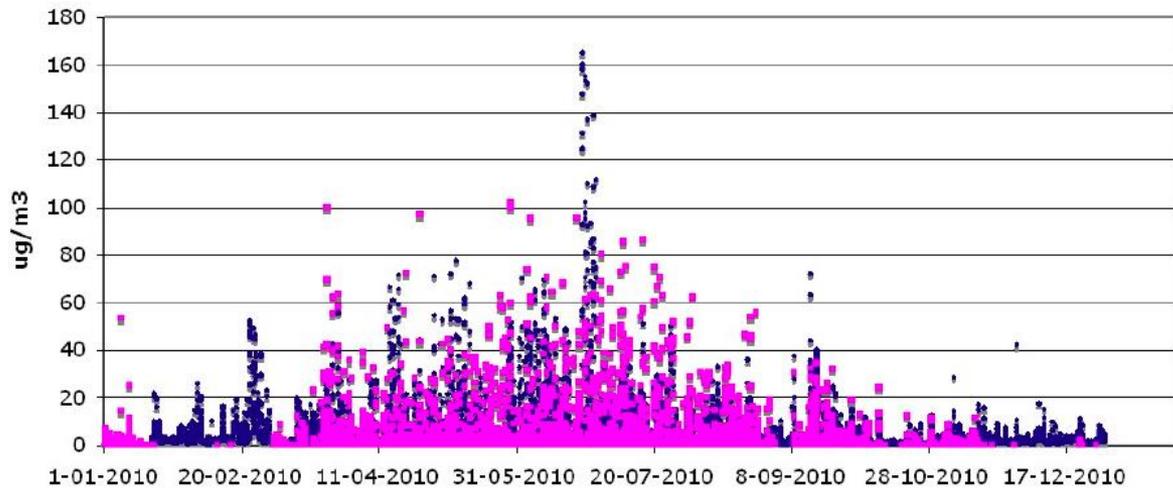
Mytilus galloprovincialis



10 ppb -
Regolamento
Europeo
1881/2006



Concentrazioni orarie nel 2010



VALORE DI
RIFERIMENTO
per il benzene nell'aria
ambiente, MEDIA
ANNUA 5 ug/m³

Figura 3: Rilevazioni orarie del benzene

I pallini blu e fucsia indicano
dati orari per due siti diversi

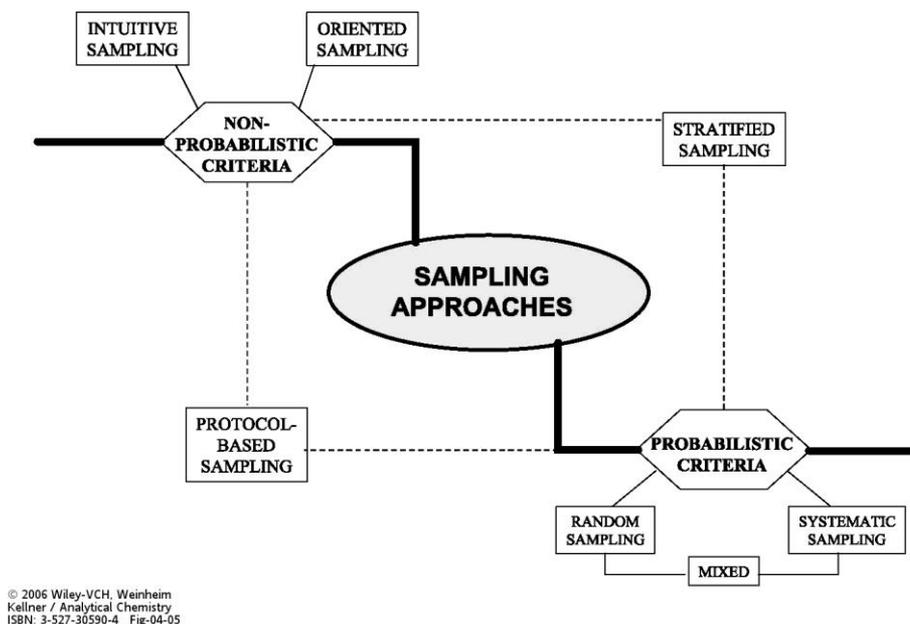


Figura 10: Direzioni da cui provengono i venti nei casi in cui le concentrazioni di benzene sono maggiori di 25 $\mu\text{g}/\text{m}^3$, nelle stazioni di via Pitacco, RFI e San Sabba (dall'alto al basso) per il 2010; il raggio del settore circolare è proporzionale alla percentuale di valori maggiori di 25 $\mu\text{g}/\text{m}^3$ che provengono dalla direzione considerata

➤ *Procedura di campionamento*

La procedura di campionamento deve essere costruita rispondendo alle seguenti domande:

- **Dove?**
- **Quando?**
- **Quanti?**
- **Che quantità?**
- **Come?**



Gli approcci della procedura di campionamento possono essere diversi:

- **basati su criteri probabilistici:** si considerano dei metodi statistici per stabilire delle regole per campionare parti del materiale (oggetto), es. campionamento di un terreno;
- **basati su criteri non-probabilistici:** in questo caso si assume che ci sia probabilità zero di poter prelevare alcune o molte parti del materiale, es. campionamento di un corso d'acqua

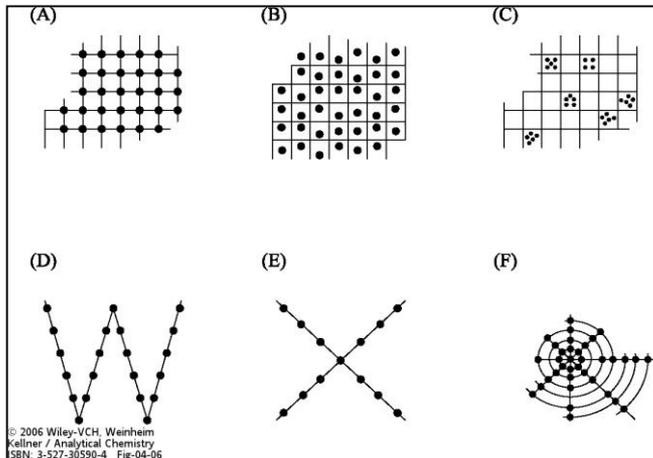
segue →

❖ Criteri probabilistici

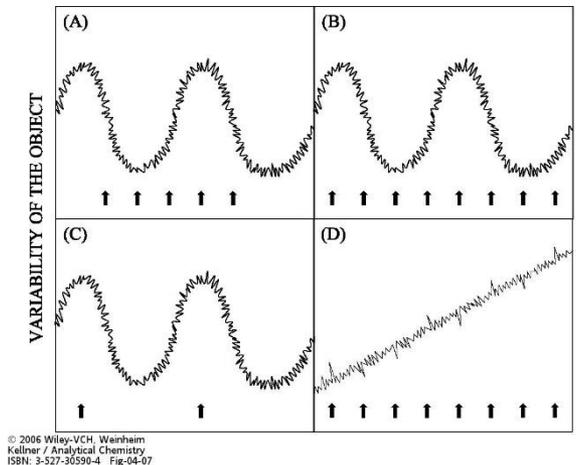
Ci sono tre tipi di criteri probabilistici applicabili:

- a) **random**: viene condotto in modo che tutte le parti del materiale (detto "lotto") hanno la stessa probabilità di essere campionate. Ciò di solito comporta la raccolta di un grande numero di campioni;
- b₁) **sistematico nello spazio**: i campioni vengono raccolti secondo specifici schemi spaziali scelti a seconda delle caratteristiche del materiale e delle informazioni che si vogliono ricavare. Questo è il metodo più applicato;
- b₂) **sistematico nel tempo**: quando l'oggetto dell'indagine è di natura dinamica (cioè la sua composizione cambia in maniera ciclica o meno nel tempo) si possono scegliere diversi intervalli temporali a seconda degli scopi;
- c) **misto**: per sistemi molto complessi si possono applicare approcci misti.

Approccio
sistematico
nello spazio



Approccio
sistematico
nel tempo



segue →

❖ **Criteri non-probabilistici**

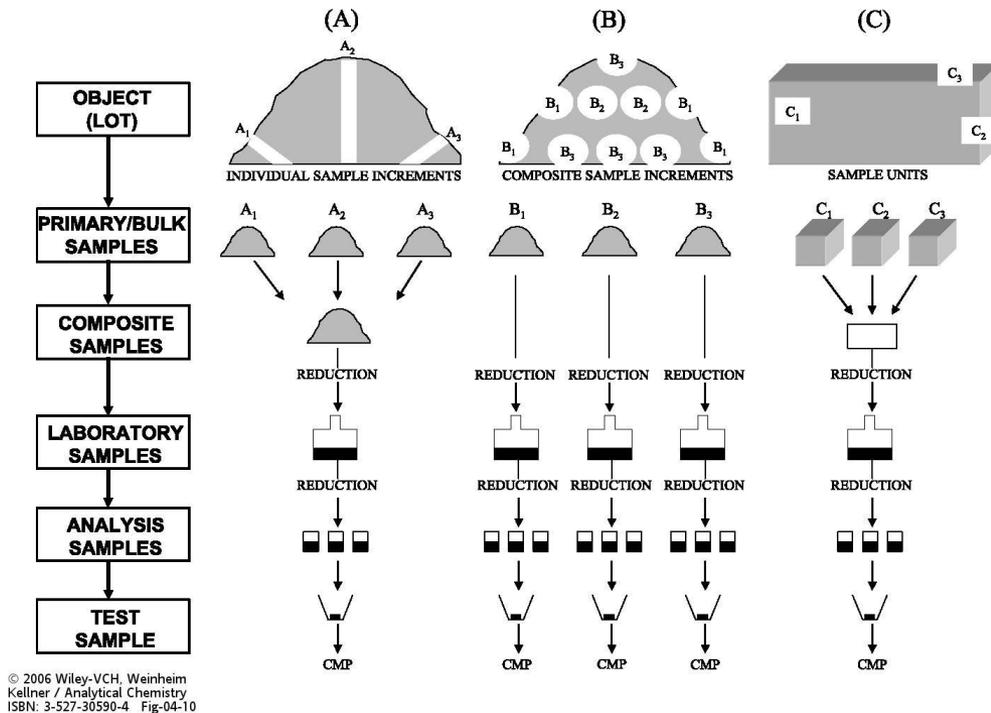
Ci sono due tipi di criteri non-probabilistici applicabili:

- a) **intuitivo**: l'analista, in base alla sua personale esperienza, sceglie le porzioni di materiale da analizzare, ad es. quando visivamente c'è una parte del materiale alterata di cui si vogliono determinare le caratteristiche;
- b) **orientato**: è la scelta migliore quando il problema analitico richiede informazioni molto specifiche e ben definite, ad es. la concentrazione di metalli in particelle sospese di una determinata dimensione in acque naturali.

➤ **Tipi di campione**

- ✓ Un **campione** è definito come: una **porzione** (cioè una parte singola, discreta, ben definita) prelevata dal materiale oggetto di indagine in accordo con la procedura di campionamento utilizzata;
- ✓ Il **materiale** (oggetto) può essere definito tramite il termine statistico **popolazione** (cioè un insieme finito o infinito di parti che differiscono da parti di popolazioni diverse);
- ✓ Un **lotto** è un materiale (oggetto) di cui si assume che tutte le parti appartengano alla **stessa popolazione**.

- ✓ Un **campione** può anche essere definito in base alla procedura con cui è stato prelevato (vedi slides precedenti), oppure in base alla sua **dimensione** e "**distanza**" (in termini di passaggi intermedi di riduzione di dimensione) dal materiale oggetto di indagine.



(CMP = Chemical Measurement Process)

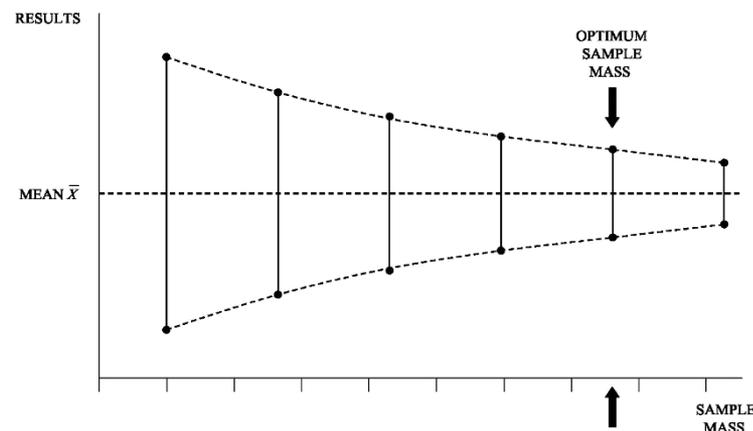
Sintesi su campionamento matrici alimentari <https://people.umass.edu/~mcclemen/581Sampling.html>

➤ Errori di campionamento

- **Errori sistematici:** sono dovuti a **difetti** della procedura di campionamento, ad es.
 - apparecchiature inappropriate,
 - conservazione del campione inadeguata,
 - aumento o diminuzione di massa dovuta ad interazioni con il contenitore di trasporto,
 - reattività con agenti atmosferici (H_2O , CO_2 , O_2 , metabolismo di microorganismi).

Possono avere segno positivo o negativo, e sono, in linea di principio, eliminabili poiché identificabili.

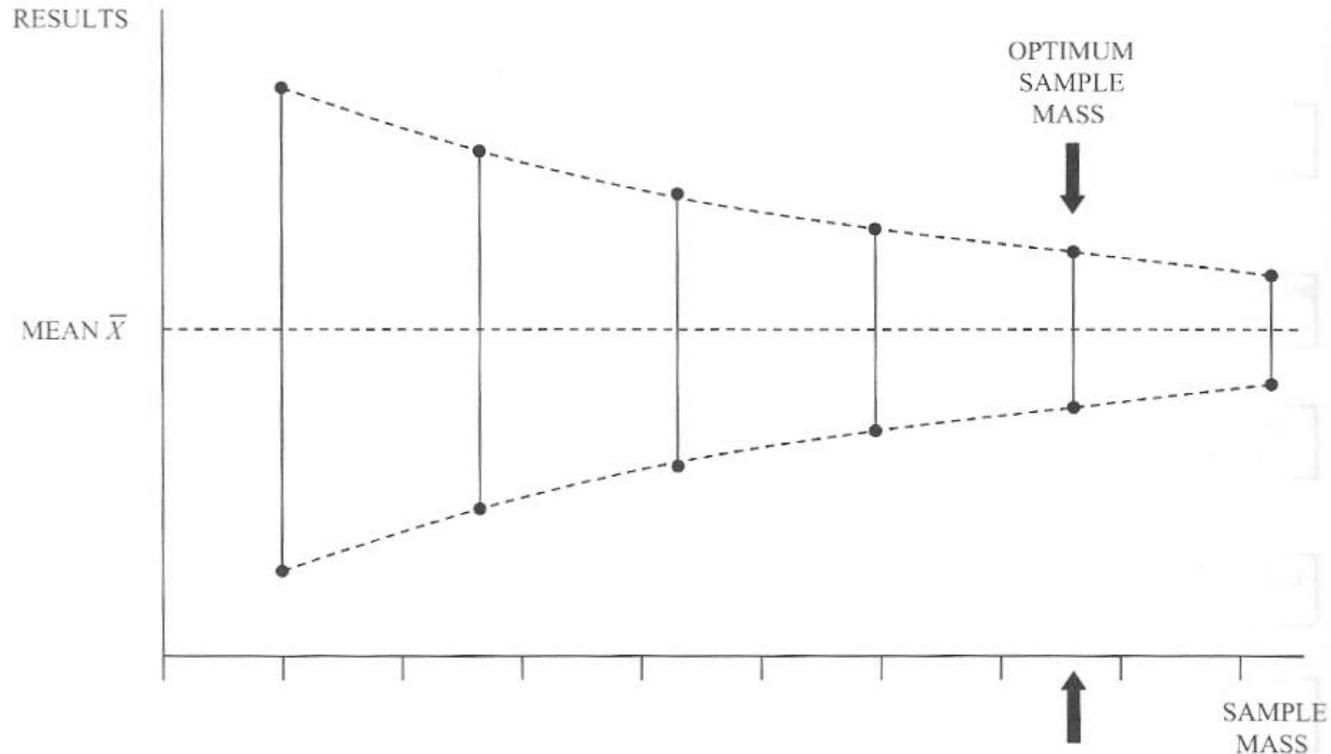
- **Errori casuali/random:** derivano dalla **variabilità** inerente la procedura di campionamento ed, in particolare, la mancaza di omogeneità del materiale (oggetto). In genere diminuiscono con l'aumentare della quantità di campione e con il numero dei campioni. Assumendo che l'eterogeneità del materiale abbia una distribuzione gaussiana, questi errori possono essere quantificati attraverso la **deviazione standard**.



© 2006 Wiley-VCH, Weinheim
Kellner / Analytical Chemistry
ISBN: 3-527-30590-4 Fig.04-11

Errore relativo in relazione alla massa di campione

Figure 4.11. Random sampling errors (represented by interval of results around the mean value (•—•)) as a function of sample mass



“Ingamells has shown that the relation $WR^2 = K_s$ is valid in many situations. W represents the weight of sample analyzed, R is the relative standard deviation (in percent) of sample composition, and K_s is the sampling constant, corresponding to the weight of sample required to limit the sampling uncertainty to 1% with 68% confidence. The magnitude of K_s may be determined by estimating s_s from a series of measurements of samples of weight W ”
Kratochvil, An.Chem. 1981 <http://pubs.acs.org/doi/pdf/10.1021/ac00231a001>

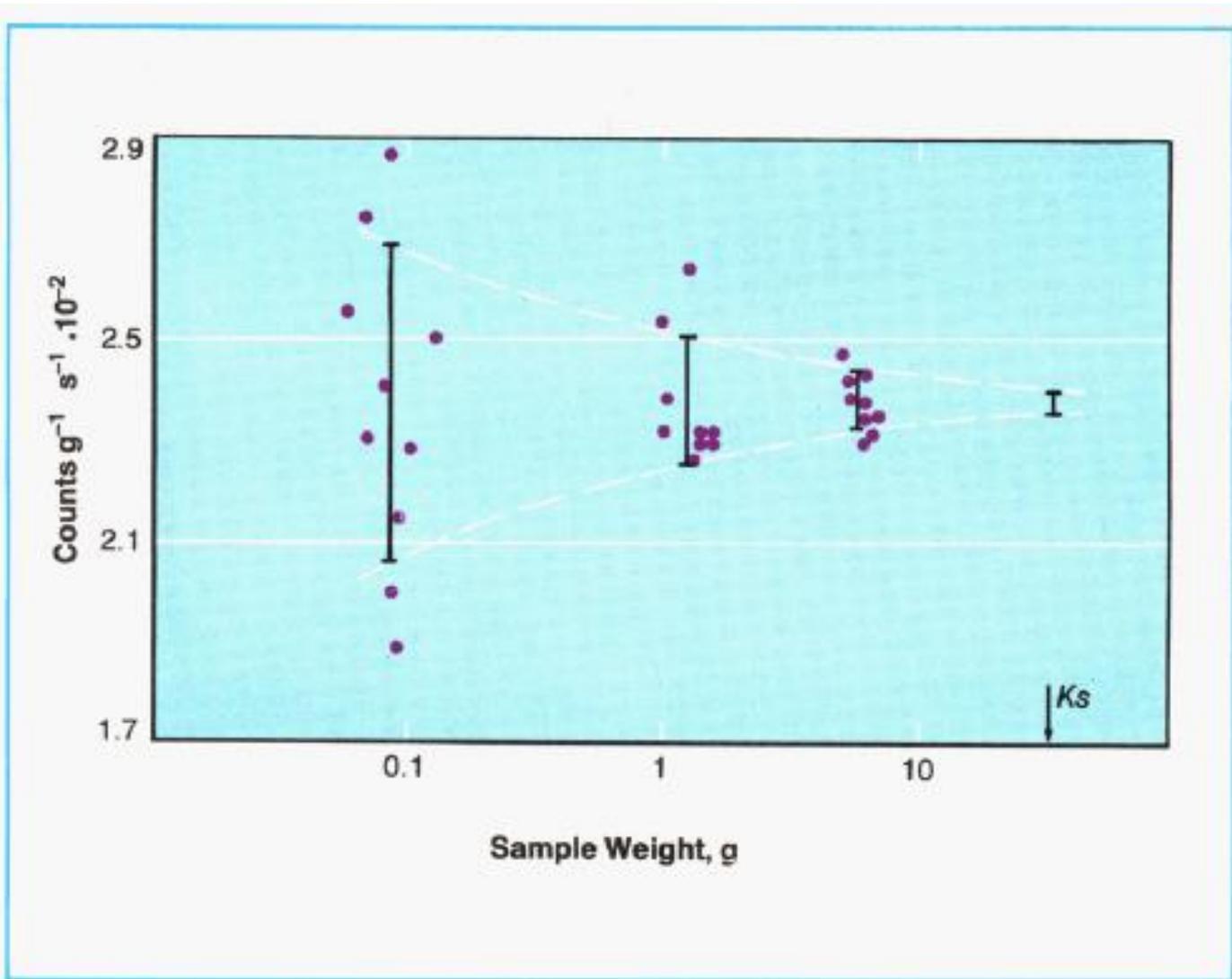


Figure 3. Sampling diagram of sodium-24 in human liver homogenate (from Reference 7)

IL SITO ACQUARIO

1998

2007

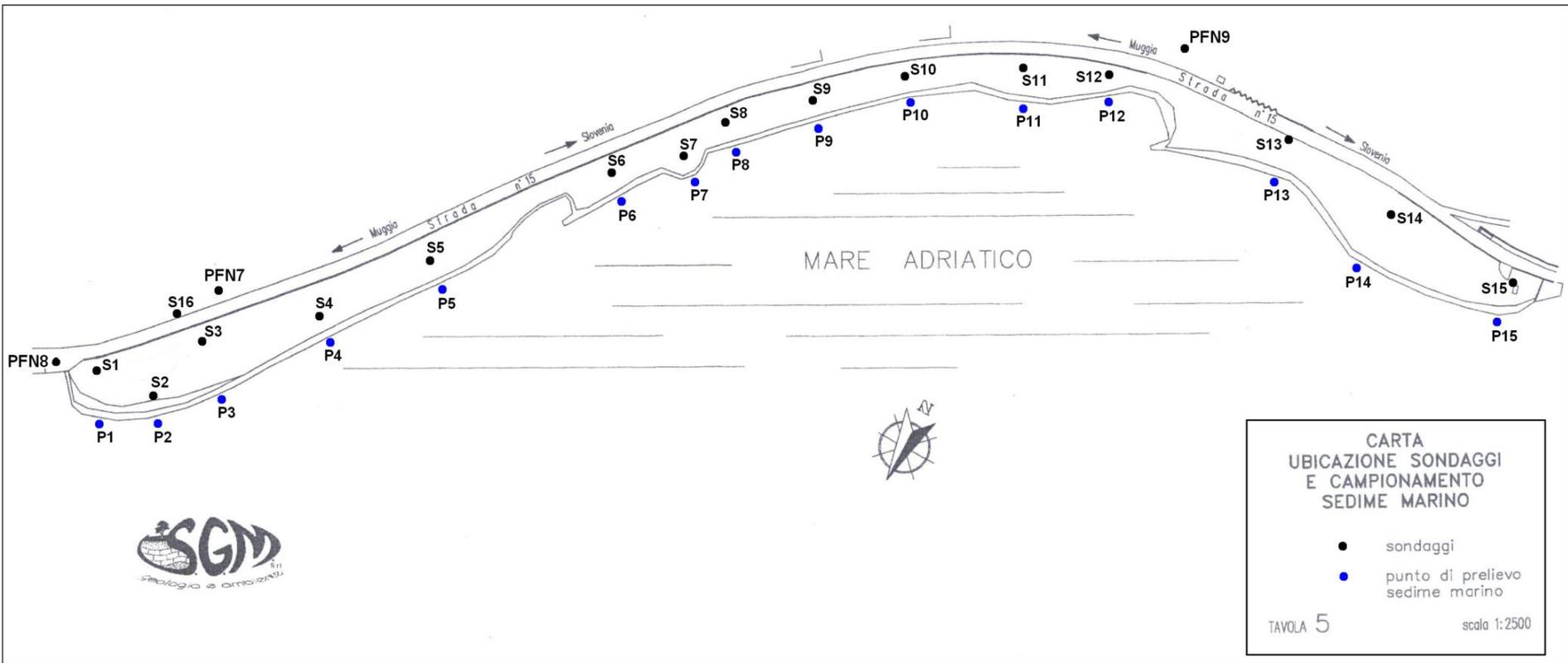


Terrapieno di colmata (fascia a mare a valle della strada Provinciale n.14 Muggia-Lazzaretto).

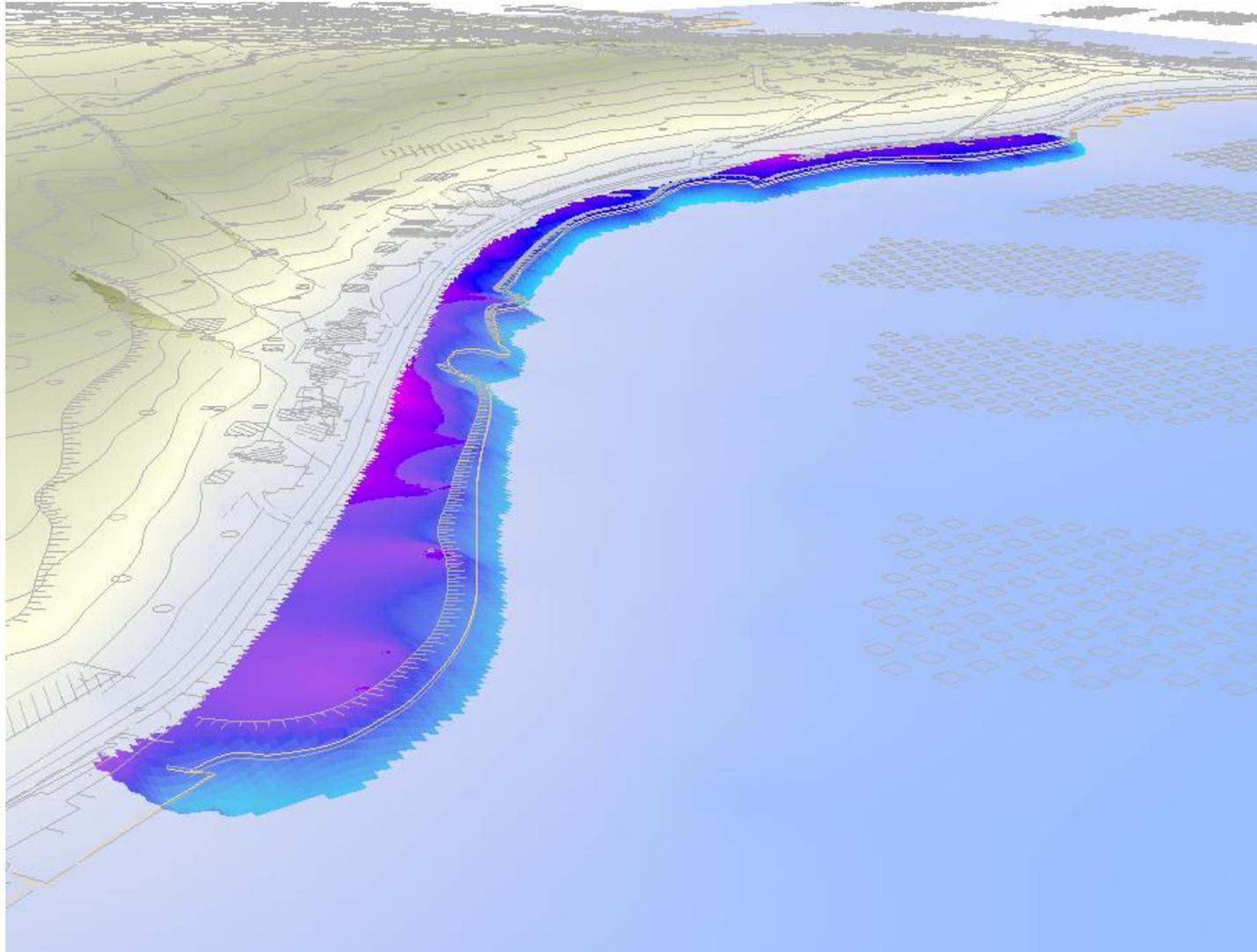
Superficie 28.094 m² = 24.583 m² di terrapieno + 3.511 m² di scogliera.

Estensione lungo-costa di circa 1.000 m - larghezza media di circa 30 m.
Considerando uno spessore medio del terreno di riporto di circa 4 m, il volume totale del terrapieno risulta di circa 100.000 m³.

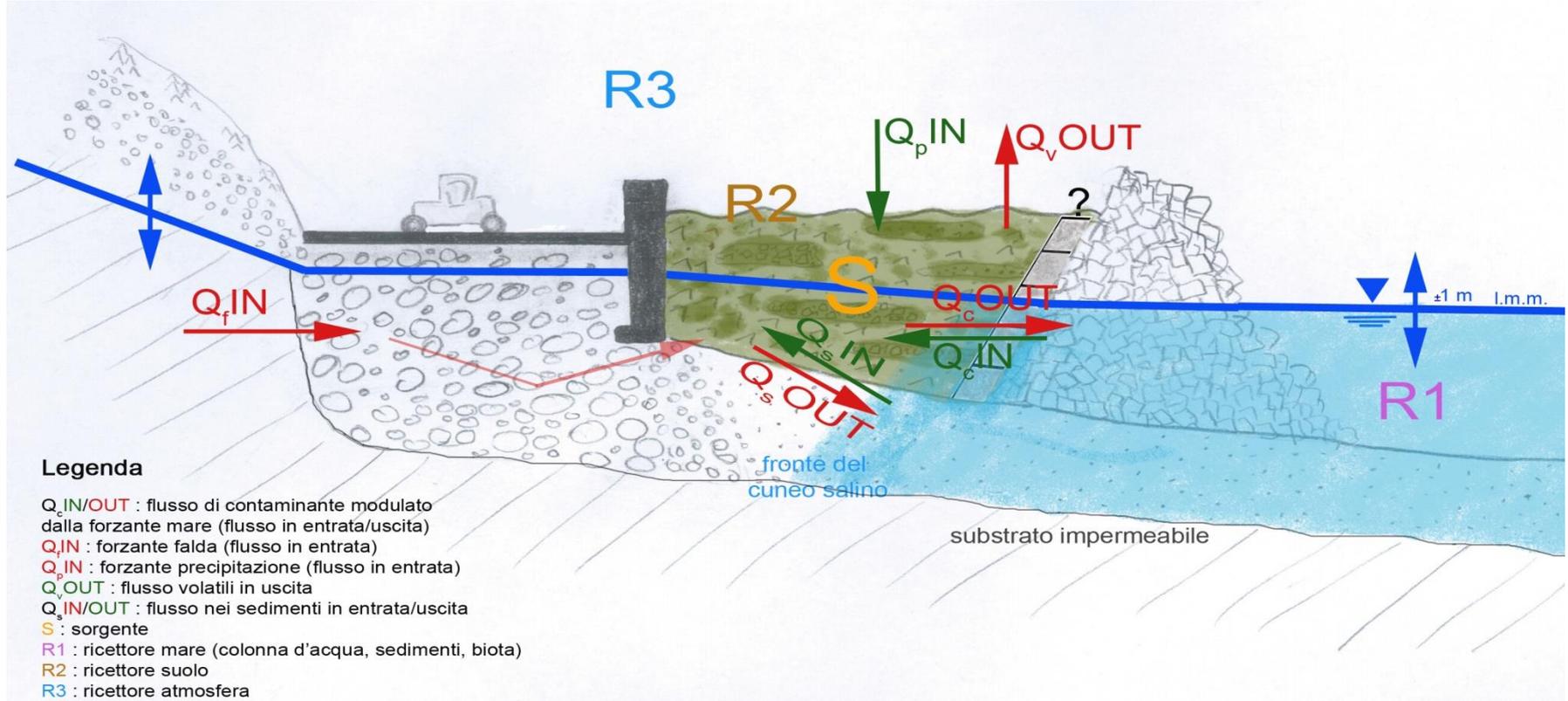
Ubicazione sondaggi e sedime marino



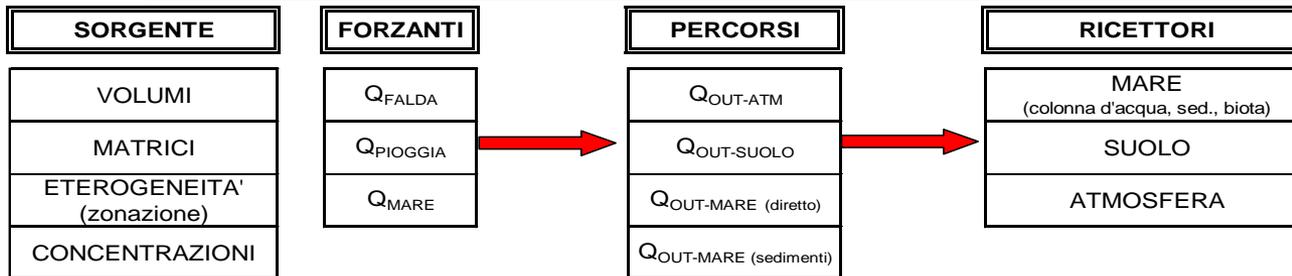
DATABASE PER ANALISI DATI E PIANIFICAZIONE INTERVENTI



Nuovo MODELLO CONCETTUALE PRELIMINARE



SORGENTE ⇒ TRASPORTO ⇒ BERSAGLI





ANALISI CHIMICHE su suoli e acque

- UNITS - DISGAM (Referente Covelli) granulometria, Umidità, Carbonio Organico, Mercurio
- UNITS - DSCh (Referente Barbieri) IPA, PCB, OT
- UNITS - DSCh (Referente Adami) Metalli
- EnviroLab d.o.o.(referente Addobbati) Amianto e C>12
- Ambiente Analisi srl (referente Sancandi) speciazione C>12 per analisi di rischio, IPA (HPLC-FD)



Introduzione alla chemiometria

Chemimetria

Matematica

Statistica

Scienze dell'Informazione

In Chimica

Discipline simili

- Biometrics ± 1900
- Psychometrics ± 1930
- Econometrics ± 1950
- Technometrics ± 1960

Qualche dato storico

- **Nome** proposto originariamente nei primi anni 1970 dal chimico organico svedese Svante Wold.
- **International Chemometrics Society** - 1970s.
- Meeting Internazionale - **Cosenza** 1983
- **Riviste** : 1986 (Chemometrics and Intelligent Laboratory Systems) and 1987 (J Chemometrics)
- **Libri** : metà anni 1980
- **Corsi** : nei tardi anni 1980 principalmente come formazione professionale continua.

Cosa è la chemiometria

La chemiometria è un settore della chimica che studia l'applicazione dei metodi matematici o statistici ai dati chimici. La International Chemometrics Society (ICS) ne dà la seguente definizione: *la chemiometria è la scienza di relazionare le misure effettuate su un sistema o su un processo chimico allo stato del sistema via applicazione di metodi matematici o statistici.*

<http://www.gruppochemiometria.it>

[Home](#)
[Chi siamo](#)
[La Chemiometria](#)
[Attività](#)
[News](#)
[Software](#)
[Contatti](#)

Cosa è la Chemiometria

La chemiometria è un settore della chimica che studia l'applicazione dei metodi matematici o statistici ai dati chimici. La International Chemometrics Society (ICS) ne dà la seguente definizione: "la chemiometria è la scienza di relazionare le misure effettuate su un sistema o su un processo chimico allo stato del sistema via applicazione di metodi matematici o statistici",

"a chemical discipline that uses mathematical and statistical methods to: design/select optimal procedures and experiments, provide maximum chemical information by analysing data, give a graphical representation of this information, in other words information aspects of chemistry"

La chemiometria può essere definita come la branca della chimica che si serve di metodi matematici, statistici e logici per:

- progettare, selezionare ed ottimizzare procedure ed esperimenti;
- estrarre la massima informazione possibile sul sistema in esame attraverso l'analisi dei dati;
- fornire una rappresentazione grafica di questa informazione.

Appare chiaro come la chemiometria accompagni il processo chimico, ed in particolare chimico-analitico, lungo tutte le sue fasi a partire dal campionamento fino all'ottimizzazione.

La programmazione relativa all'anno **2022** è la seguente:

17-21 Gennaio: **Analisi Multivariata** (il corso verrà svolto in modalità telematica)

16-20 Maggio: **Experimental Design** (il corso verrà svolto in modalità telematica)

26-30 Settembre: **Analisi Multivariata** (in inglese, il corso verrà svolto in modalità telematica)

7-11 Novembre: **Experimental Design** (in inglese, il corso verrà svolto in modalità telematica)

Le quote di partecipazione, per la scuola di Gennaio, sono le seguenti:

	Quota (€)	Quota ridotta (€)*
Industria	800	640
Istituzioni no-profit	550	440
Studenti, Dottorandi, Assegnisti,...	400	320

*Le **quote ridotte** sono previste per agevolare l'iscrizione di più partecipanti provenienti da uno stesso Dipartimento o Ditta (la quota si intende per ciascuno degli iscritti), per partecipanti che abbiano già frequentato una delle due scuole e per partecipanti dell'Ateneo Genovese.

A partire dalla scuola di Maggio le quote di partecipazione subiranno una leggera variazione.

«...La spettroscopia NIR è ormai riconosciuta come una delle tecniche più versatili ed efficaci nell'ambito dell'**analisi rapida e non distruttiva** di alimenti, matrici ambientali, di prodotti/manufatti industriali e artigianali, intermedi di processo ecc. In particolare, per i **suoi costi contenuti, la possibilità di essere implementata at-/on-/in-line, la vasta gamma di soluzioni portatili e miniaturizzate** è diventata la **tecnica di riferimento in molti ambiti per il controllo di qualità ed il monitoraggio di processo**. La regione spettrale NIR comprende gli **assorbimenti dovuti agli overtones (transizioni da numero quantico vibrazionale $v=0$ a $v=2$) e alle bande di combinazione**, è pertanto una **regione molto sensibile alle modificazioni dell'intorno chimico in cui sono inseriti i legami/gruppi funzionali che danno luogo ad assorbimenti**, d'altro canto è anche caratterizzata da **bande fortemente sovrapposte e variabilità dovuta alle caratteristiche fisiche del campione e alle condizioni di misura**; per estrarre informazioni utili dai profili spettrali NIR, acquisiti su matrici complesse, è quasi **imprescindibile l'uso di metodi chemiometrici** (analisi multivariata dei segnali)...»

Marina Cocchi, Tiziana Cattaneo «Focus sulla spettroscopia NIR»

https://www.soc.chim.it/sites/default/files/chimind/pdf/2014_9_3733_on.pdf

[https://chem.libretexts.org/Textbook_Maps/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_\(Physical_and_Theoretical_Chemistry\)/Spectroscopy/Vibrational_Spectroscopy/Vibrational_Modes/Combination_Bands%2C_Overtones_and_Fermi_Resonances](https://chem.libretexts.org/Textbook_Maps/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Spectroscopy/Vibrational_Spectroscopy/Vibrational_Modes/Combination_Bands%2C_Overtones_and_Fermi_Resonances)

"A chemical discipline that uses mathematical and statistical methods to: design/select optimal procedures and experiments, provide maximum chemical information by analysing data, give a graphical representation of this information, in other words... information aspects of chemistry" (D.L.Massart)

La chemiometria può essere definita come la branca della chimica che si serve di metodi matematici, statistici e logici per:

- progettare, selezionare ed ottimizzare procedure ed esperimenti;
- estrarre la massima informazione possibile sul sistema in esame attraverso l'analisi dei dati;
- fornire una rappresentazione grafica di questa informazione.

(modificato da

<http://www.iupac.org/publications/pac/pdf/1983/pdf/5512x1861.pdf>)

Appare chiaro come la chemiometria accompagni il processo chimico, ed in particolare chimico-analitico, lungo tutte le sue fasi a partire dal campionamento fino all'ottimizzazione.

Campi di applicazione della chemiometria

Tra i campi d'applicazione della chemiometria si possono citare:

- Controllo di qualità
- Monitoraggio e controllo di processo
- Tracciabilità degli alimenti
- QSAR/QSPR e REACH
- Genomica, proteomica e metabolomica
- Progettazione degli esperimenti e Ottimizzazione
- Progettazione di Farmaci e materiali (*Drug & material design*)
- Analisi di immagini
- Applicazioni in ambito industriale e ambientale

Analisi Multivariata dei Dati

- Dati campionati e progetti con molte risposte anche da:
 - Attività minerarie
 - Ospedali
 - Agricoltura
 - Industria alimentare
 - Etc.

La Chemiometria è una disciplina per l'analisi dei dati che:

- Tratta dati **multivariati** (e “**multiway**”)
- Si basa su modellizzazione **soft**
- Usa **metodi di proiezione** e il concetto di **variabili latenti**
- Considera i **dati** come **informazione + rumore**
- Considera il **rumore** come **informazione inutile**

Nomenclatura

- I campioni sono **oggetti**
- Ciò che è misurato su un oggetto è una **variabile**

Dati multivariati

Variabili

Campioni

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15
s1	-1.19E-01	7.28E-01	-2.15E-02	5.22E-01	7.06E-04	7.32E-01	3.10E-04	-6.13E-04	-5.92E-05	1.28E+00	1.32E+00	-7.03E-02	1.23E-03	4.67E-01	-4.21E-01
s2	-1.37E-01	7.28E-01	-2.89E-02	6.08E-01	7.09E-04	7.02E-01	6.58E-04	-1.22E-03	-1.49E-04	1.35E+00	1.39E+00	-3.27E-01	2.48E-04	4.84E-01	-2.94E-01
s3	2.51E-02	-9.15E-02	6.73E-03	-1.13E-01	-9.07E-05	-7.58E-02	-2.29E-04	4.10E-04	5.65E-05	-1.96E-01	-2.02E-01	1.49E-01	3.83E-04	-6.80E-02	1.43E-01
s4	-1.14E-01	6.70E-01	-2.18E-02	5.04E-01	6.50E-04	6.65E-01	3.83E-04	-7.34E-04	-7.96E-05	1.20E+00	1.24E+00	-1.36E-01	8.59E-04	4.34E-01	-1.09E-01
s5	-7.93E-02	4.14E-01	-1.69E-02	3.51E-01	4.04E-04	3.98E-01	3.96E-04	-7.35E-04	-9.05E-05	7.71E-01	7.94E-01	-2.02E-01	7.80E-05	2.76E-01	-1.83E-01
s6	1.51E-02	-6.38E-02	3.74E-03	-6.75E-02	-6.28E-05	-5.67E-02	-1.15E-04	2.07E-04	2.78E-05	-1.29E-01	-1.33E-01	7.08E-02	1.40E-04	-4.52E-02	6.71E-01
s7	7.44E-02	-5.24E-01	1.11E-02	-3.24E-01	-5.06E-04	-5.45E-01	-1.73E-05	7.92E-05	-1.07E-05	-8.87E-01	-9.13E-01	-1.02E-01	-1.47E-03	-3.26E-01	-1.19E-01
s8	3.65E-02	-2.66E-01	5.12E-03	-1.59E-01	-2.56E-04	-2.78E-01	1.43E-05	-3.95E-07	-1.14E-05	-4.46E-01	-4.59E-01	-6.86E-02	-8.12E-04	-1.64E-01	-7.71E-01
s9	1.36E-01	-7.06E-01	2.89E-02	-6.01E-01	-6.88E-04	-6.77E-01	-6.83E-04	1.26E-03	1.56E-04	-1.31E+00	-1.35E+00	3.50E-01	-1.12E-04	-4.71E-01	3.18E-01
s10	-2.74E-02	3.60E-01	1.82E-03	1.12E-01	3.42E-04	4.12E-01	-4.31E-04	7.24E-04	1.22E-04	5.29E-01	5.43E-01	3.97E-01	2.27E-03	2.02E-01	4.03E-01
s11	7.47E-02	-3.31E-01	1.80E-02	-3.34E-01	-3.25E-04	-2.99E-01	-5.30E-04	9.62E-04	1.28E-04	-6.54E-01	-6.74E-01	3.20E-01	5.44E-04	-2.31E-01	3.01E-01
s12	-1.17E-01	7.02E-01	-2.16E-02	5.13E-01	6.81E-04	7.03E-01	3.40E-04	-6.63E-04	-6.76E-05	1.25E+00	1.28E+00	-9.79E-02	1.07E-03	4.52E-01	-7.02E-01
s13	1.06E-01	-2.82E-01	3.23E-02	-4.82E-01	-2.85E-04	-1.87E-01	-1.25E-03	2.21E-03	3.14E-04	-7.01E-01	-7.25E-01	8.59E-01	2.71E-03	-2.36E-01	8.33E-01
s14	7.39E-02	-5.28E-01	1.07E-02	-3.21E-01	-5.09E-04	-5.50E-01	2.49E-06	4.48E-05	-1.59E-05	-8.90E-01	-9.15E-01	-1.17E-01	-1.54E-03	-3.27E-01	-1.34E-01
s15	-9.87E-03	1.02E-01	-3.21E-04	4.17E-02	9.75E-05	1.13E-01	-8.29E-05	1.36E-04	2.44E-05	1.57E-01	1.61E-01	8.35E-02	5.31E-04	5.92E-02	8.57E-01
s16	-1.06E-01	7.68E-01	-1.52E-02	4.62E-01	7.41E-04	8.03E-01	-2.24E-05	-2.68E-05	2.88E-05	1.29E+00	1.33E+00	1.86E-01	2.30E-03	4.75E-01	2.11E-01
s17	-4.76E-02	2.66E-01	-9.52E-03	2.10E-01	2.59E-04	2.61E-01	1.92E-04	-3.61E-04	-4.19E-05	4.84E-01	4.99E-01	-8.35E-02	2.29E-04	1.75E-01	-7.21E-01
s18	9.54E-02	-6.55E-01	1.48E-02	-4.16E-01	-6.33E-04	-6.77E-01	-6.69E-05	1.79E-04	-1.61E-06	-1.12E+00	-1.15E+00	-9.35E-02	-1.71E-03	-4.10E-01	-1.16E-01
s19	-1.32E-01	5.01E-01	-3.49E-02	5.94E-01	4.96E-04	4.22E-01	1.16E-03	-2.09E-03	-2.86E-04	1.06E+00	1.09E+00	-7.49E-01	-1.84E-03	3.68E-01	-7.17E-01
s20	8.91E-02	-4.23E-01	2.05E-02	-3.97E-01	-4.14E-04	-3.94E-01	-5.56E-04	1.02E-03	1.32E-04	-8.15E-01	-8.40E-01	3.19E-01	3.49E-04	-2.90E-01	2.97E-01
s21	-8.91E-02	5.08E-01	-1.75E-02	3.93E-01	4.94E-04	5.01E-01	3.34E-04	-6.35E-04	-7.18E-05	9.17E-01	9.45E-01	-1.36E-01	5.26E-04	3.31E-01	-1.15E-01
s22	1.15E-01	-6.22E-01	2.39E-02	-5.10E-01	-6.06E-04	-6.04E-01	-5.25E-04	9.79E-04	1.18E-04	-1.14E+00	-1.18E+00	2.52E-01	-3.18E-04	-4.12E-01	2.24E-01
s23	-4.08E-02	5.43E-01	2.94E-03	1.67E-01	5.17E-04	6.22E-01	-6.60E-04	1.11E-03	1.86E-04	7.96E-01	8.18E-01	6.06E-01	3.45E-03	3.04E-01	6.15E-01
s24	9.92E-02	-6.00E-01	1.82E-02	-4.36E-01	-5.82E-04	-6.02E-01	-2.81E-04	5.49E-04	5.53E-05	-1.06E+00	-1.09E+00	7.64E-02	-9.46E-04	-3.86E-01	5.28E-01
s25	1.08E-01	-5.37E-01	2.40E-02	4.81E-01	-5.25E-04	-5.07E-01	-6.15E-04	1.13E-03	1.43E-04	-1.02E+00	-1.05E+00	3.37E-01	1.85E-04	-3.63E-01	3.11E-01
s26	-6.95E-02	4.56E-01	-1.15E-02	3.04E-01	4.41E-04	4.67E-01	1.03E-04	-2.23E-04	-1.34E-05	7.88E-01	8.11E-01	2.38E-02	1.03E-03	2.88E-01	4.01E-01
s27	4.90E-02	-1.23E-01	1.51E-02	-2.22E-01	-1.25E-04	-7.71E-02	-5.94E-04	1.05E-03	1.50E-04	-3.15E-01	-3.25E-01	4.12E-01	1.33E-03	-1.05E-01	4.00E-01
s28	-1.65E-03	-7.79E-02	-3.41E-03	1.09E-02	-7.26E-05	-1.01E-01	2.35E-04	-4.05E-04	-6.30E-05	-8.88E-02	-9.08E-02	-1.91E-01	-8.92E-04	-3.68E-02	-1.90E-01
s29	8.73E-02	-5.70E-01	1.46E-02	-3.82E-01	-5.52E-04	-5.83E-01	-1.37E-04	2.94E-04	1.91E-05	-9.87E-01	-1.02E+00	-2.30E-02	-1.27E-03	-3.61E-01	-4.36E-01
s30	-6.93E-02	6.98E-02	-2.51E-02	3.19E-01	7.94E-05	-2.22E-02	1.11E-03	-1.96E-03	-2.85E-04	3.19E-01	3.31E-01	-8.06E-01	-2.95E-03	9.79E-02	-7.88E-01
s31	-8.99E-02	3.66E-01	-2.28E-02	4.03E-01	3.61E-04	3.20E-01	7.21E-04	-1.30E-03	-1.76E-04	7.48E-01	7.72E-01	-4.52E-01	-9.80E-04	2.63E-01	-4.30E-01
s32	-6.32E-02	2.05E-01	-1.79E-02	2.85E-01	2.05E-04	1.59E-01	6.44E-04	-1.15E-03	-1.61E-04	4.63E-01	4.78E-01	-4.31E-01	-1.23E-03	1.59E-01	-4.15E-01
s33	-1.42E-01	6.98E-01	-3.20E-02	6.33E-01	6.83E-04	6.57E-01	8.32E-04	-1.53E-03	-1.95E-04	1.33E+00	1.37E+00	-4.62E-01	-3.32E-04	4.74E-01	-4.28E-01
s34	1.32E-01	-6.89E-01	2.81E-02	-5.85E-01	-6.72E-04	-6.62E-01	-6.61E-04	1.23E-03	1.51E-04	-1.28E+00	-1.32E+00	3.38E-01	-1.27E-04	-4.60E-01	3.06E-01
s35	-1.08E-01	4.80E-01	-2.61E-02	4.84E-01	4.72E-04	4.35E-01	7.66E-04	-1.39E-03	-1.84E-04	9.49E-01	9.79E-01	-4.61E-01	-7.76E-04	3.36E-01	-4.34E-01
s36	2.13E-02	-2.11E-01	9.91E-04	-9.02E-02	-2.02E-04	-2.34E-01	1.57E-04	-2.55E-04	-4.66E-05	-3.28E-01	-3.38E-01	-1.62E-01	-1.06E-03	-1.24E-01	-1.67E-01
s37	-2.39E-03	4.55E-03	-7.90E-04	1.09E-02	4.75E-06	1.94E-03	3.28E-05	-5.80E-05	-8.35E-06	1.36E-02	1.41E-02	-2.32E-02	-7.97E-05	4.43E-03	-2.26E-01
s38	6.29E-02	-3.26E-01	1.35E-02	-2.79E-01	-3.18E-04	-3.13E-01	-3.21E-04	5.94E-04	7.36E-05	-6.09E-01	-6.27E-01	1.66E-01	-3.74E-05	-2.18E-01	1.51E-01
s39	1.02E-01	-5.03E-01	2.26E-02	-4.51E-01	-4.91E-04	-4.74E-01	-5.80E-04	1.07E-03	1.35E-04	-9.53E-01	-9.83E-01	3.19E-01	1.85E-04	-3.40E-01	2.94E-01
s40	1.00E-01	-4.22E-01	2.50E-02	-4.49E-01	-4.16E-04	-3.75E-01	-7.68E-04	1.39E-03	1.86E-04	-8.52E-01	-8.79E-01	4.75E-01	9.52E-04	-3.00E-01	4.51E-01
s41	7.05E-02	-3.24E-01	1.66E-02	-3.14E-01	-3.17E-04	-2.97E-01	-4.70E-04	8.56E-04	1.12E-04	-6.31E-01	-6.51E-01	2.77E-01	3.94E-04	-2.24E-01	2.60E-01
s42	-7.27E-02	3.21E-01	-1.76E-02	3.24E-01	3.15E-04	2.90E-01	5.17E-04	-9.38E-04	-1.25E-04	6.35E-01	6.55E-01	-3.12E-01	-5.35E-04	2.25E-01	-2.94E-01
s43	-3.34E-02	2.26E-01	-5.31E-03	1.46E-01	2.18E-04	2.33E-01	3.24E-05	-7.79E-05	-1.84E-06	3.87E-01	3.98E-01	2.54E-02	5.64E-04	1.42E-01	3.33E-01
s44	-1.36E-01	6.25E-01	-3.21E-02	6.08E-01	6.13E-04	5.73E-01	9.11E-04	-1.66E-03	-2.18E-04	1.22E+00	1.26E+00	-5.37E-01	-7.70E-04	4.32E-01	-5.04E-01
s46	2.41E-02	-1.53E-01	4.17E-03	-1.06E-01	-1.48E-04	-1.55E-01	-4.92E-05	1.01E-04	8.29E-06	-2.67E-01	-2.75E-01	2.90E-03	-3.05E-04	-9.74E-02	-2.79E-01
s47	-9.11E-02	4.29E-01	-2.11E-02	4.06E-01	4.21E-04	3.98E-01	5.77E-04	-1.05E-03	-1.37E-04	8.29E-01	8.55E-01	-3.33E-01	-3.91E-04	2.95E-01	-3.11E-01
s49	5.84E-02	-2.37E-01	1.49E-02	-2.62E-01	-2.34E-04	-2.07E-01	-4.71E-04	8.49E-04	1.15E-04	-4.85E-01	-5.00E-01	2.96E-01	6.47E-04	-1.70E-01	2.82E-01
s50	-4.05E-02	2.15E-01	-8.50E-03	1.79E-01	2.10E-04	2.08E-01	1.93E-04	-3.58E-04	-4.36E-05	3.98E-01	4.10E-01	-9.53E-02	7.91E-05	1.43E-01	-8.56E-01

dati multivariati (e "multiway")⁴⁰

1 x
1
0D

x
 i
1
1D

X
 i
 j
2D

K
 X
 i
 j
3D

Molti inputs inducono un effetto

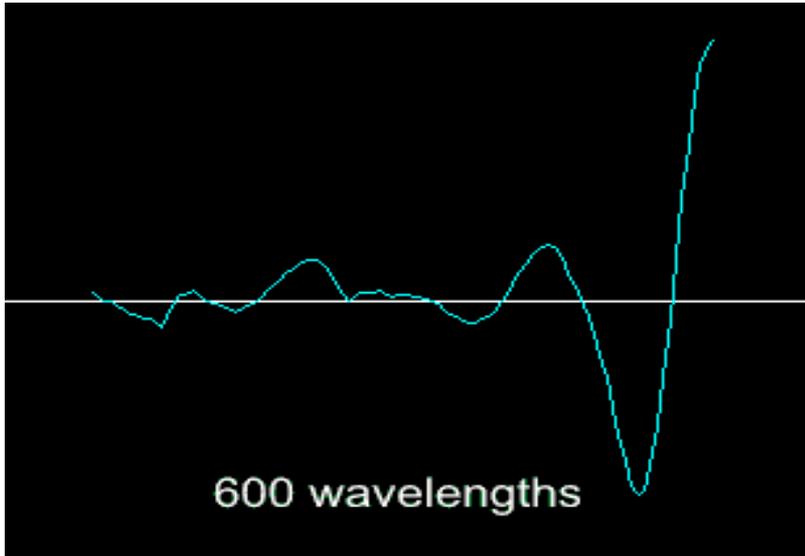
Molti effetti sono derivati da un input

etc

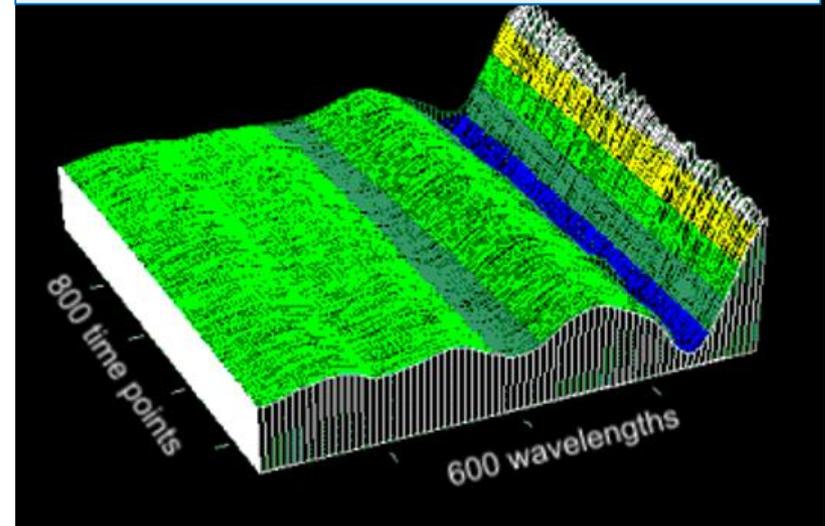
dati multivariati (e “multiway”)⁴¹

Molte variabili e molti campioni

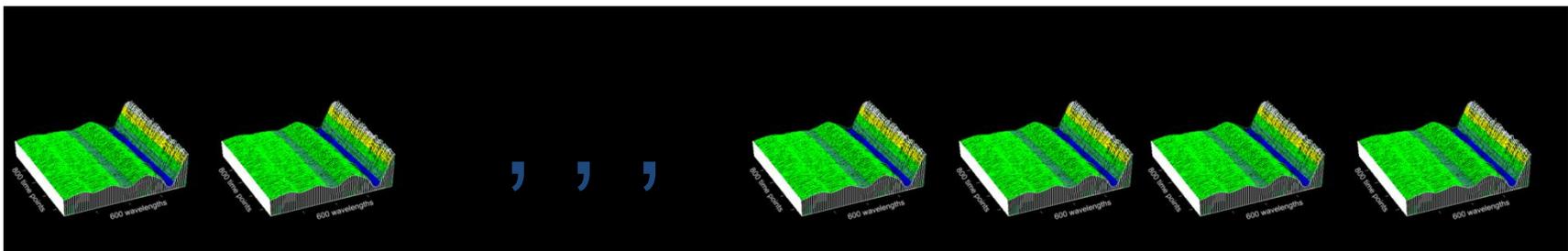
Una misura – spettro (600 punti)



Un “batch” – 800 spettri (suddivisioni temporali)



Un set di dati – 200 campioni (batches)



dati multivariati (e “multiway”)⁴²

34.92

Spettro

K

C
a
m
p
i
o
n
i

1

1

I

Vettori

12.0
3.6
11.1
5.9
34.0
0.5
1.4
17.0

Un vettore è una raccolta di numeri.

E' sempre un vettore colonna

12.0 3.6 11.1 5.9 34.0 0.5 1.4 17.0

La trasposta di un vettore è un vettore riga.

I simboli per indicare la trasposizione sono ' o T .

Es. \mathbf{a}' o \mathbf{a}^T .

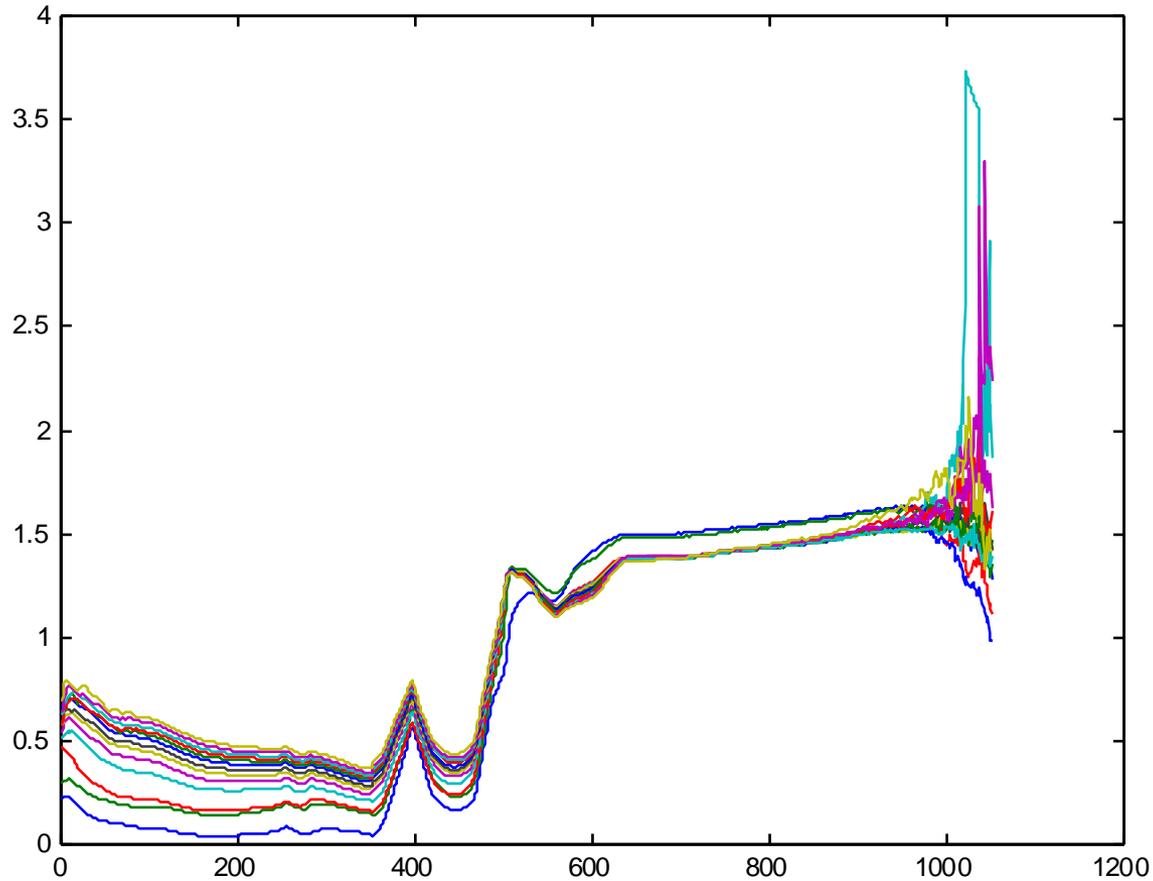
La matrice dei dati

K

Una matrice di dati
è un vettore di vettori

I

Tempi in una reazione batch



Lunghezze d'onda NIR

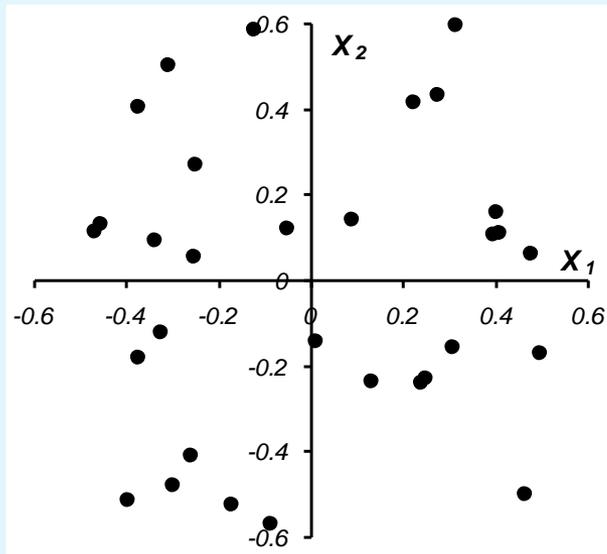
Modelli “hard” e “soft”

	Modelli “hard”	Modelli “soft”
Origine	Da conoscenza <i>a priori</i>	Dai dati
Formula	$\mathbf{y}=f(\mathbf{x},\mathbf{a})+\varepsilon$	$\mathbf{y}=\mathbf{X}\mathbf{a}+\varepsilon$
Parametri	Hanno significato fisico esplicito	Non c'è significato fisico “esplicito”
Problema	formulazione di modelli	Analisi dei dati
Scopo	estrapolazione	interpolazione
Example	Beer-Lambert	ANOVA

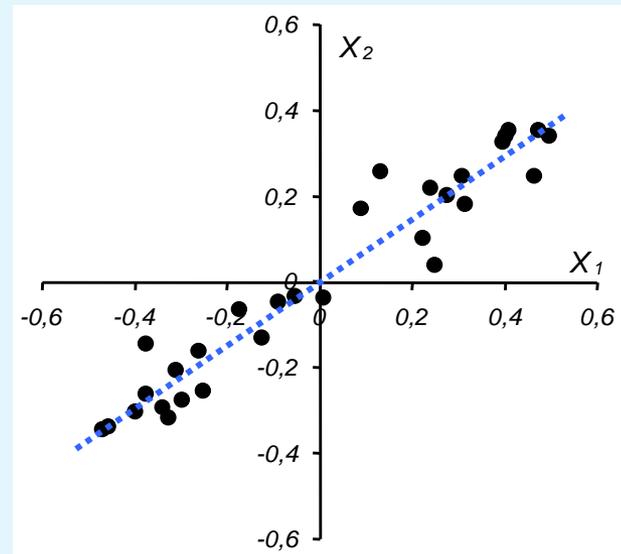
Es. L'**analisi della varianza (ANOVA)** è un insieme di tecniche [statistiche](#) che permettono di confrontare due o più gruppi di dati confrontando la variabilità *interna* a questi gruppi con la variabilità *tra* i gruppi. [Si confrontano medie di due o più campioni tenendo conto contemporaneamente di più variabili.](#)

Metodi di proiezione e variabili latenti

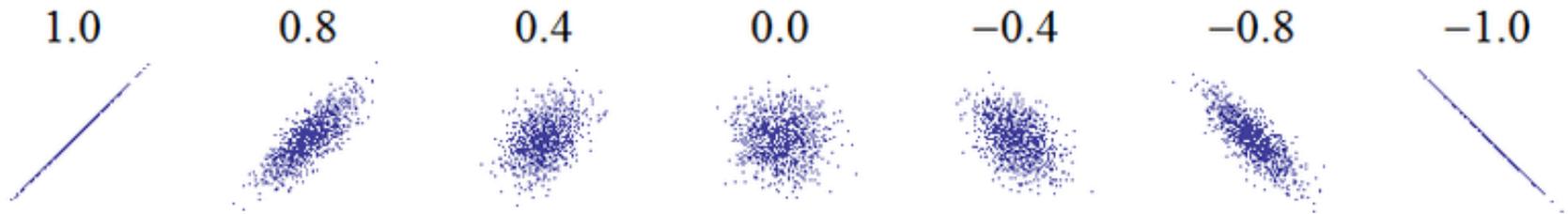
Dati senza struttura



Dati con una struttura nascosta



In [statistica](#) per **correlazione** si intende una [relazione](#) tra due [variabili](#) tale che a ciascun valore della prima [variabile](#) corrisponda con una certa regolarità un valore della seconda. Non si tratta necessariamente di un [rapporto](#) di causa ed effetto, ma semplicemente della tendenza di una variabile a variare in funzione di un'altra.



$$-1 \leq \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^n (x_i - \mu_x)^2} \sqrt{\sum_{i=1}^n (y_i - \mu_y)^2}} \leq +1$$

Grandi "aree" della Chemiometria

1. Progettazione degli esperimenti (DOE - *Design of Experiments*)
2. Analisi esplorativa dei dati (EDA - *Exploratory Data Analysis*)
3. Classificazione (*Classification*)
4. Regressione e calibrazione (*Regression and Calibration*)

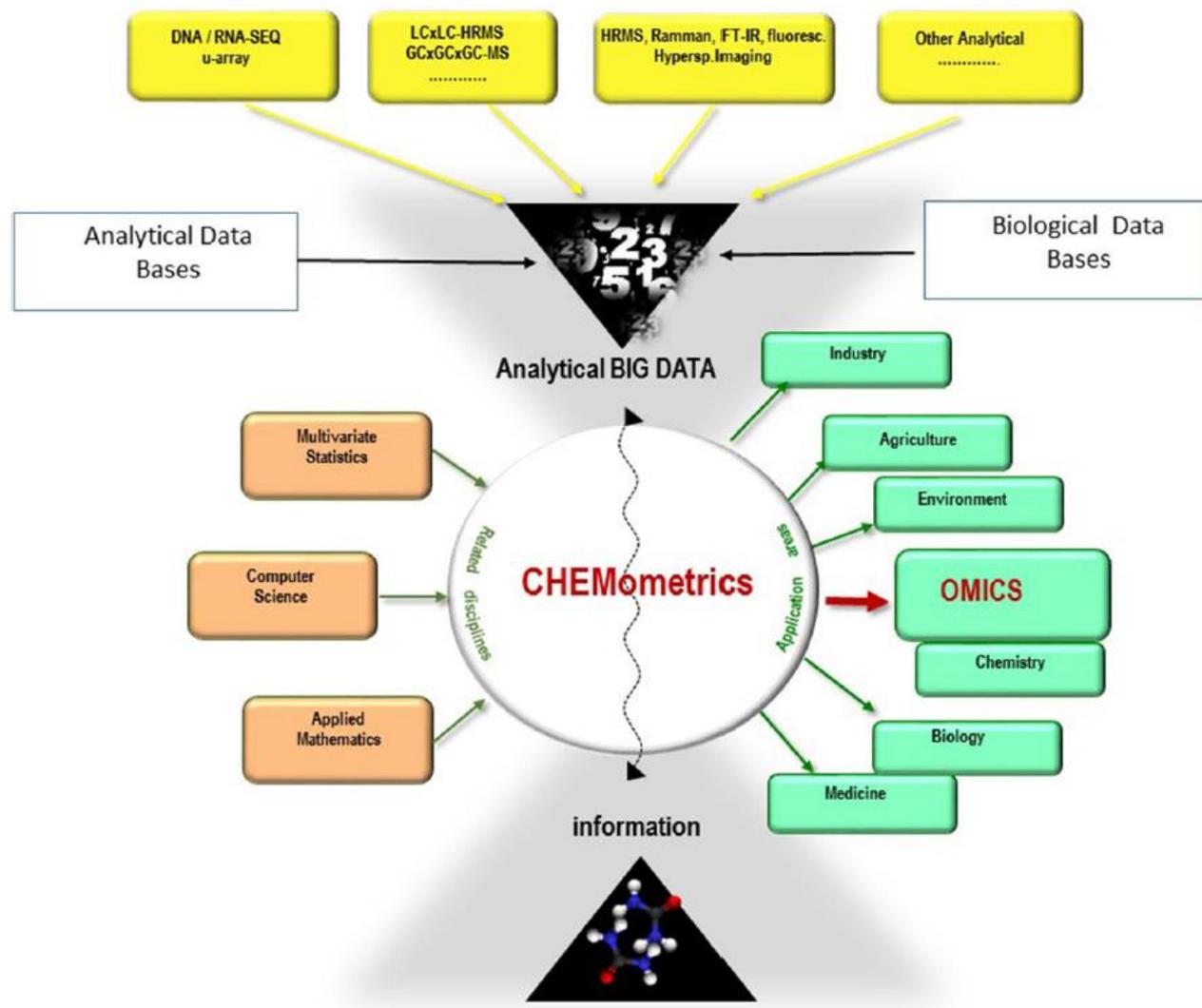
In ciascuna "area" ci sono molti metodi

Brereton et al. 2017, Chemometrics in analytical chemistry—part I: history, experimental design and data analysis tools

http://rcs.chemometrics.ru/papers/ABC409_5891.pdf

Brereton et al. 2018, Chemometrics in analytical chemistry—part II: modeling, validation, and applications DOI: 10.1007/s00216-018-1283-4

Fig. 1 Chemometrics as an interdisciplinary field



Brereton, 2018

1) Progettazione degli esperimenti

(Design of Experiments o Experimental Design)

Di estrema importanza, da applicare ove possibile

Impiega:

- ANOVA
- F-test
- t-test
- Diagrammi
- Superfici di risposta

RECUPERARE E LEGGERE

Riccardo Leardi

«Experimental design in chemistry: A tutorial»

Analytica Chimica Acta Volume 652, Issues 1-2, 2009, Pages 161-172

Progettazione degli esperimenti

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_Kx_K + b_{11}x_1^2 + b_{22}x_2^2 + \dots + b_{KK}x_K^2 + b_{12}x_1x_2 + \dots + \varepsilon$$

I Fattori x_1, x_2, \dots, x_K possono essere modificati sistematicamente

La Risposta y è misurata e modellata

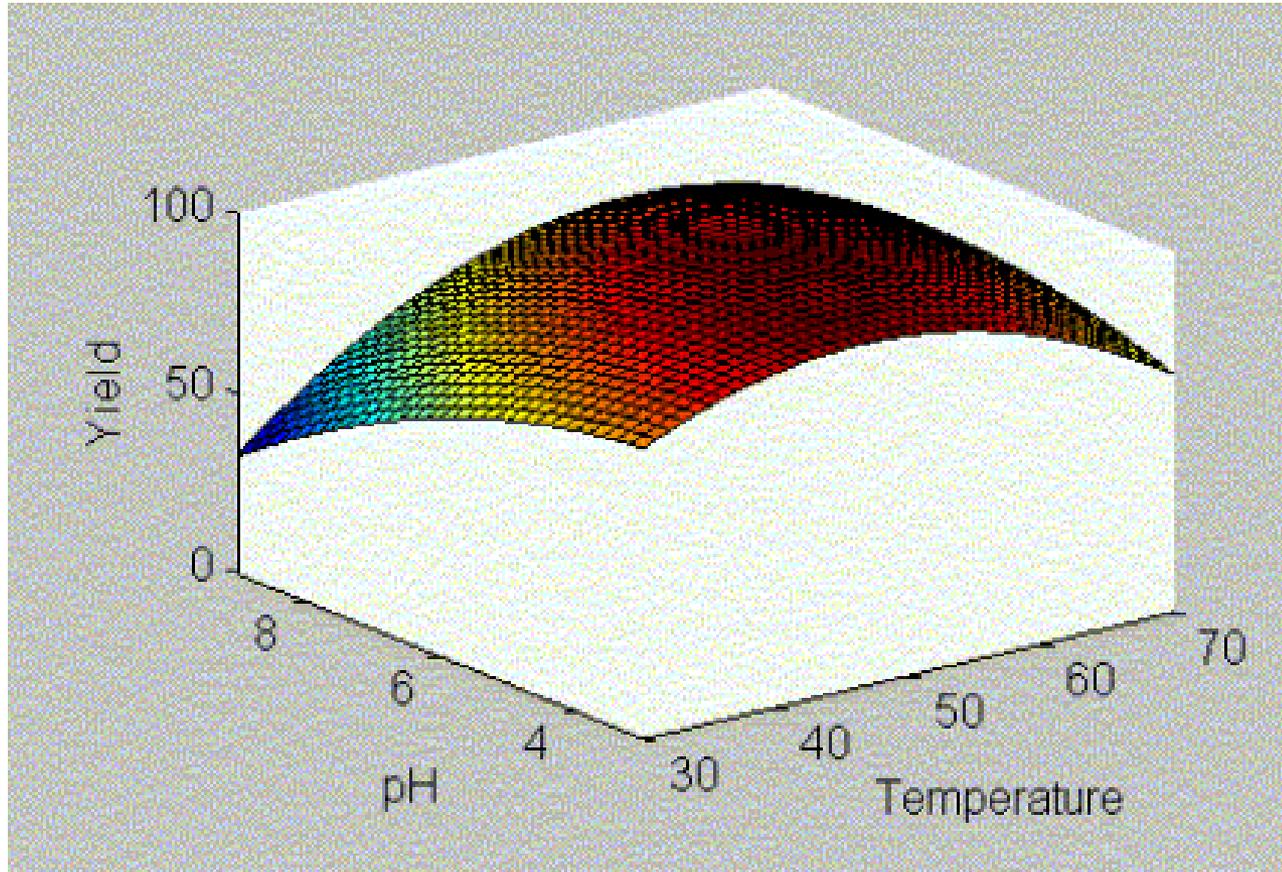
Perché progettare gli esperimenti

- *Screening* (per capire quali sono le variabili importanti nel determinare il valore di una risposta)
- *Saving time* (per risparmiare tempo)
- *Quantitative modelling* (per costruire un modello quantitativo dell'esperimento)
- *Optimisation* (per massimizzare rese di reazione, ottimizzare tempi, consumo di reagenti ...)

Perché progettare gli esperimenti?

Un problema : Ottimizzazione di una resa di reazione con pH e temperatura.

Possiamo trovare la combinazione di pH e temperatura che producono la resa migliore della reazione?

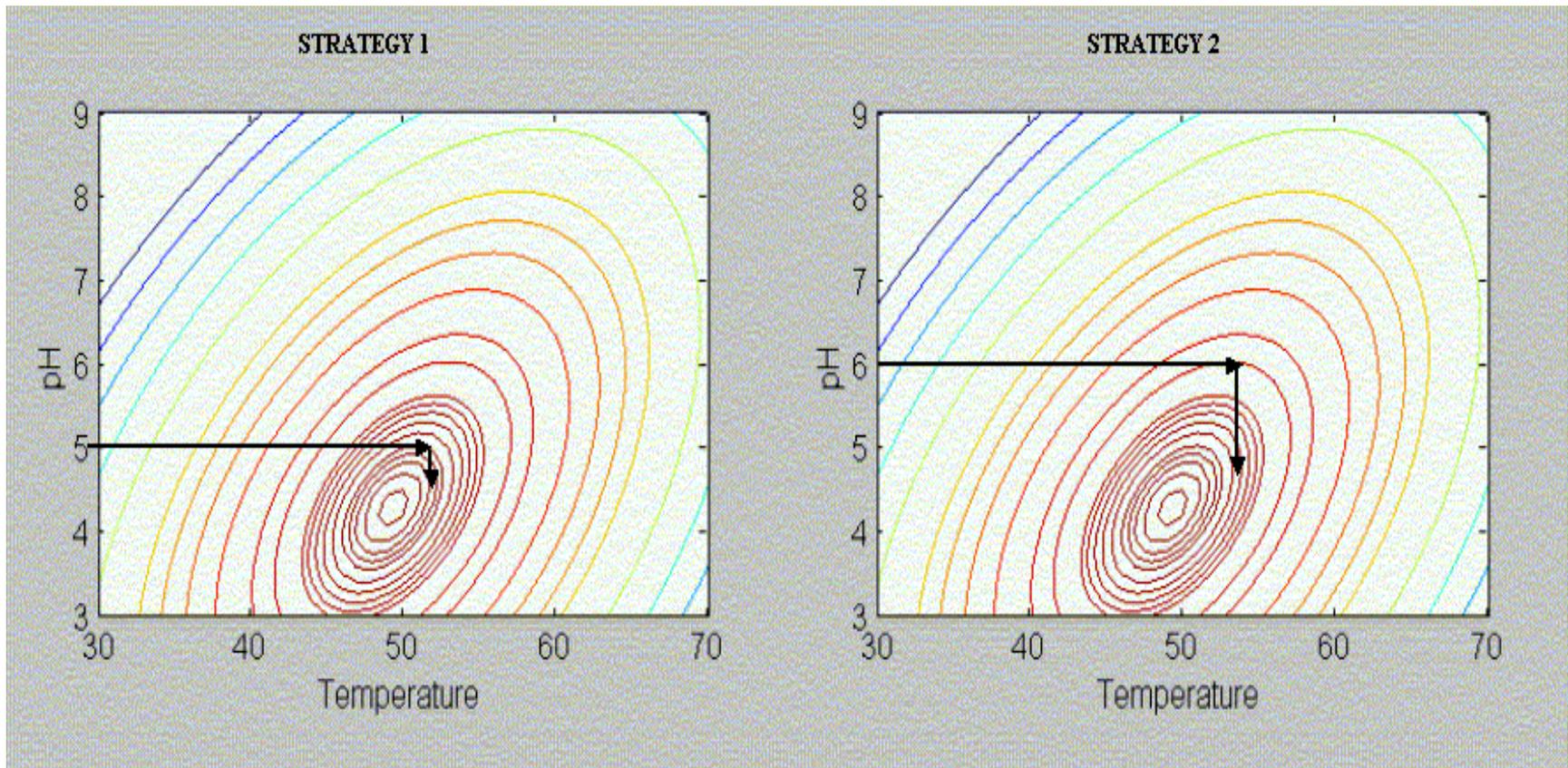


Progettazione degli esperimenti

La strategia di variare un fattore alla volta

(One Variable At Time):

può mancare di cogliere l' "ottimo"



Progettazione degli esperimenti

DIFFICOLTA'

Interazioni – la risposta per ciascun fattore non è indipendente

La temperatura ottimale a pH 5 è diversa da quella a pH 6.

Come affrontare il problema? Forza bruta?

- Una griglia di esperimenti (*Grid search*). 10 pHs, 10 temperatures, 100 experiments.
- Si inizia con una griglia a maglia larga. Poi a maglia più stretta.

Controindicazioni

- **Dispendioso in termini di tempo e denaro.**
- **Molti esperimenti vengono condotti in aree del dominio sperimentale che sono quasi sicuramente “non vicine” a un ottimo (quindi una perdita di tempo e denaro)**
- **Come stimare riproducibilità ed errore sperimentale? (Altri esperimenti, replica dei precedenti ?!?)**

Che facciamo?

Abbiamo bisogno di regole !

La progettazione formale degli esperimenti

[Analytica Chimica Acta Volume 652, Issues 1-2,](#)

12 October 2009, Pages 161-172

Experimental design in chemistry: A tutorial

[Riccardo Leardi](#)

https://www.academia.edu/237865/Experimental_Design_in_Chemistry_a_Tutorial

Progettazione degli esperimenti

Screening

- Factorial designs
- Partial factorials and Plackett-Burman designs

Modelling and optimisation

- Response surface designs
- Mixture designs



Divisione di Chimica Analitica
SOCIETÀ CHIMICA ITALIANA



GRUPPO DIVISIONALE DI CHEMIOMETRIA



UNIVERSITA' DEGLI STUDI DI GENOVA

Research Group of Analytical Chemistry and Chemometrics, of the Department of Pharmacy of the University of Genoa organizes a SCHOOL OF EXPERIMENTAL DESIGN (May 27-31, 2019).

Program:

27 May 10:00 – 13:00 (optional): basics of univariate statistics and introduction to multivariate analysis (PCA).

27 May 14:00 – 31 May 13:00: Full Factorial Designs, Screening Designs, Fractional Factorial Designs, Response Surface Methodology (Central Composite Design and Doehlert Design), “Multicriteria Decision Making”, D-Optimal Designs, Designs with qualitative variables having more than two levels, Mixture Designs, Mixture-Process Designs.

The school will be made by theoretical lessons and example illustrations, with hands-on-computer sessions in an R-based free software.

The number of seats is limited. **Contact us to know the availability before paying the registration fees.**

La programmazione relativa all'anno 2022 è la seguente:

17-21 Gennaio: Analisi Multivariata (il corso verrà svolto in modalità telematica)

16-20 Maggio: Experimental Design (il corso verrà svolto in modalità telematica)

26-30 Settembre: Analisi Multivariata (in inglese, il corso verrà svolto in modalità telematica)

7-11 Novembre: Experimental Design (in inglese, il corso verrà svolto in modalità telematica)

Le quote di partecipazione, per la scuola di Gennaio, sono le seguenti:

	Quota (€)	Quota ridotta (€)*
Industria	800	640
Istituzioni no-profit	550	440
Studenti, Dottorandi, Assegnisti,...	400	320

*Le **quote ridotte** sono previste per agevolare l'iscrizione di più partecipanti provenienti da uno stesso Dipartimento o Ditta (la quota si intende per ciascuno degli iscritti), per partecipanti che abbiano già frequentato una delle due scuole e per partecipanti dell'Ateneo Genovese.

A partire dalla scuola di Maggio le quote di partecipazione subiranno una leggera variazione.

li esperimenti

2) **Analisi Esplorativa dei Dati**

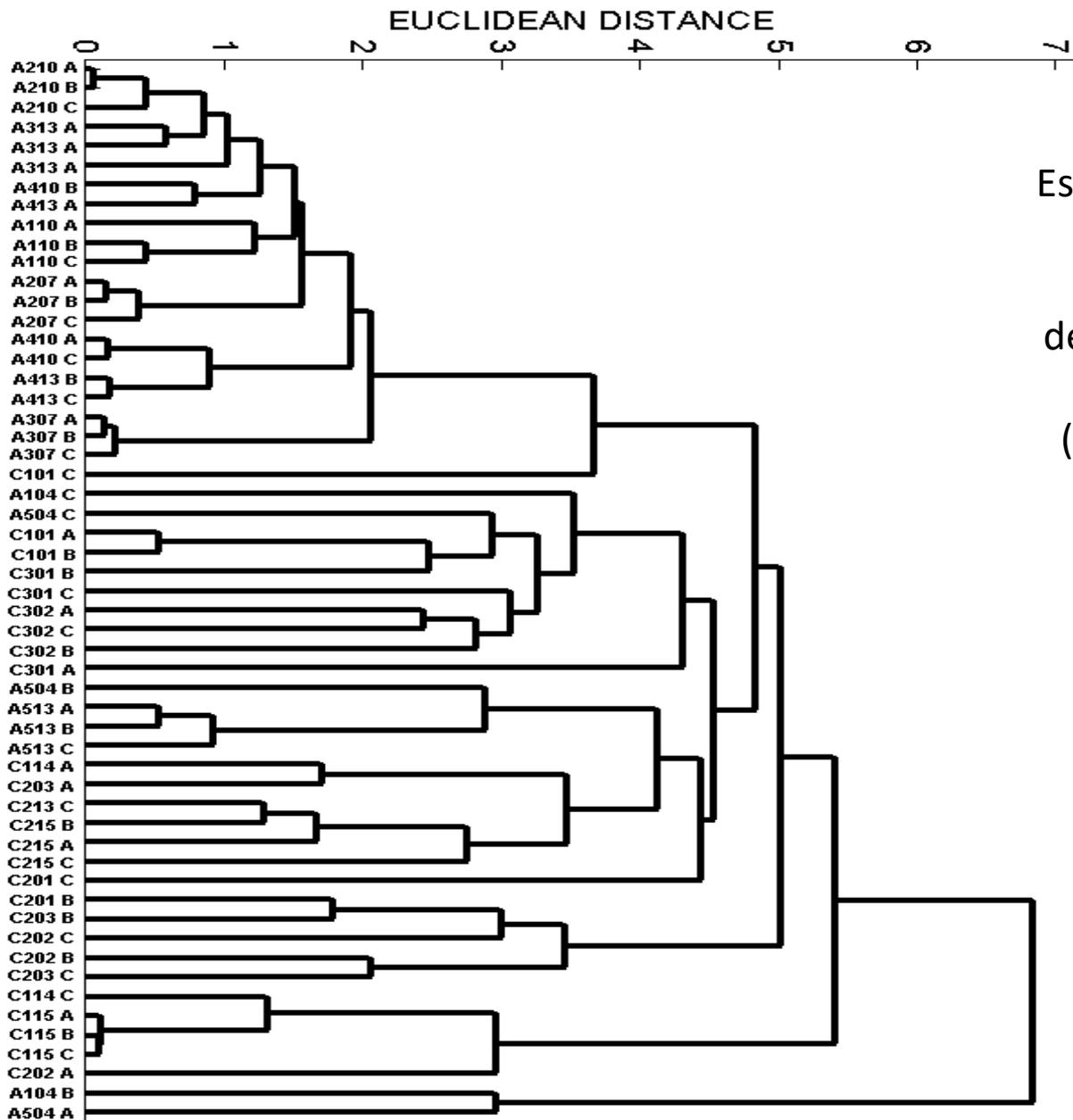
- Non ci è stato possibile progettare

Vogliamo:

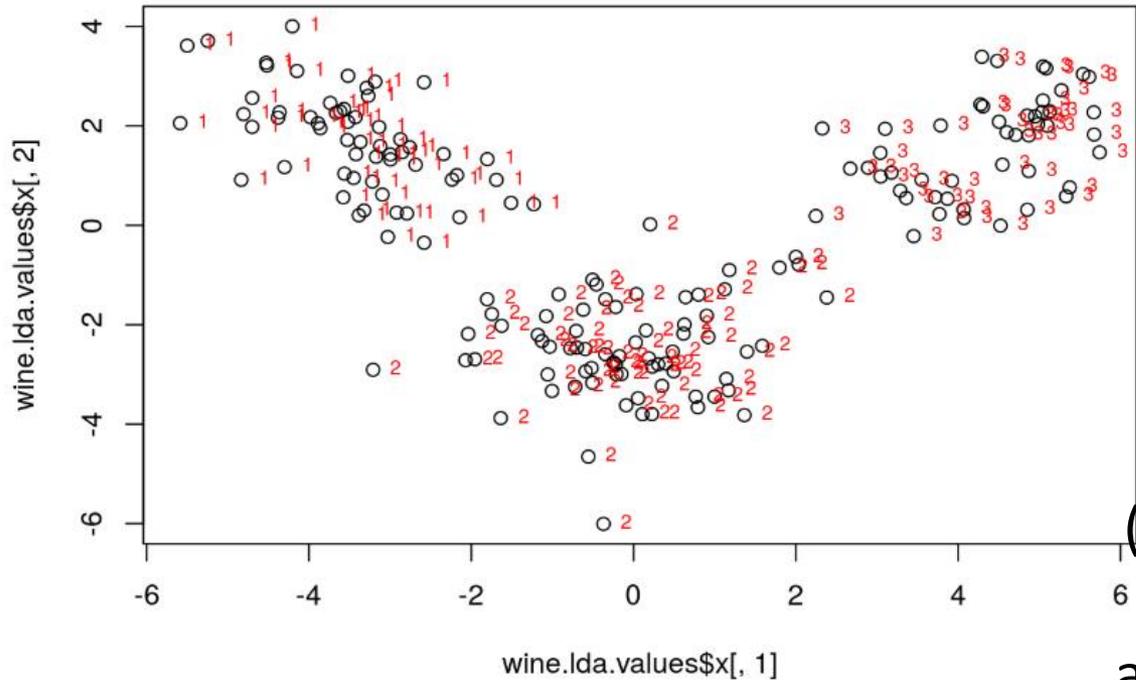
- Trovare strutture
- Trovare raggruppamenti
- Trovare dati anomali (outliers)

3) Metodi di Classificazione

- Ricerca di raggruppamenti (di campioni, molecole, etc.)= UNSUPERVISED classification
- I raggruppamenti sono noti = SUPERVISED classification
- Visualizzare i raggruppamenti
- Classificare
- Testare/validare la classificazione



Esempio di output di un'analisi di classificazione "non supervisionata": un dendrogramma di un'analisi di raggruppamento gerarchico (*hyerarchical cluster analysis*)

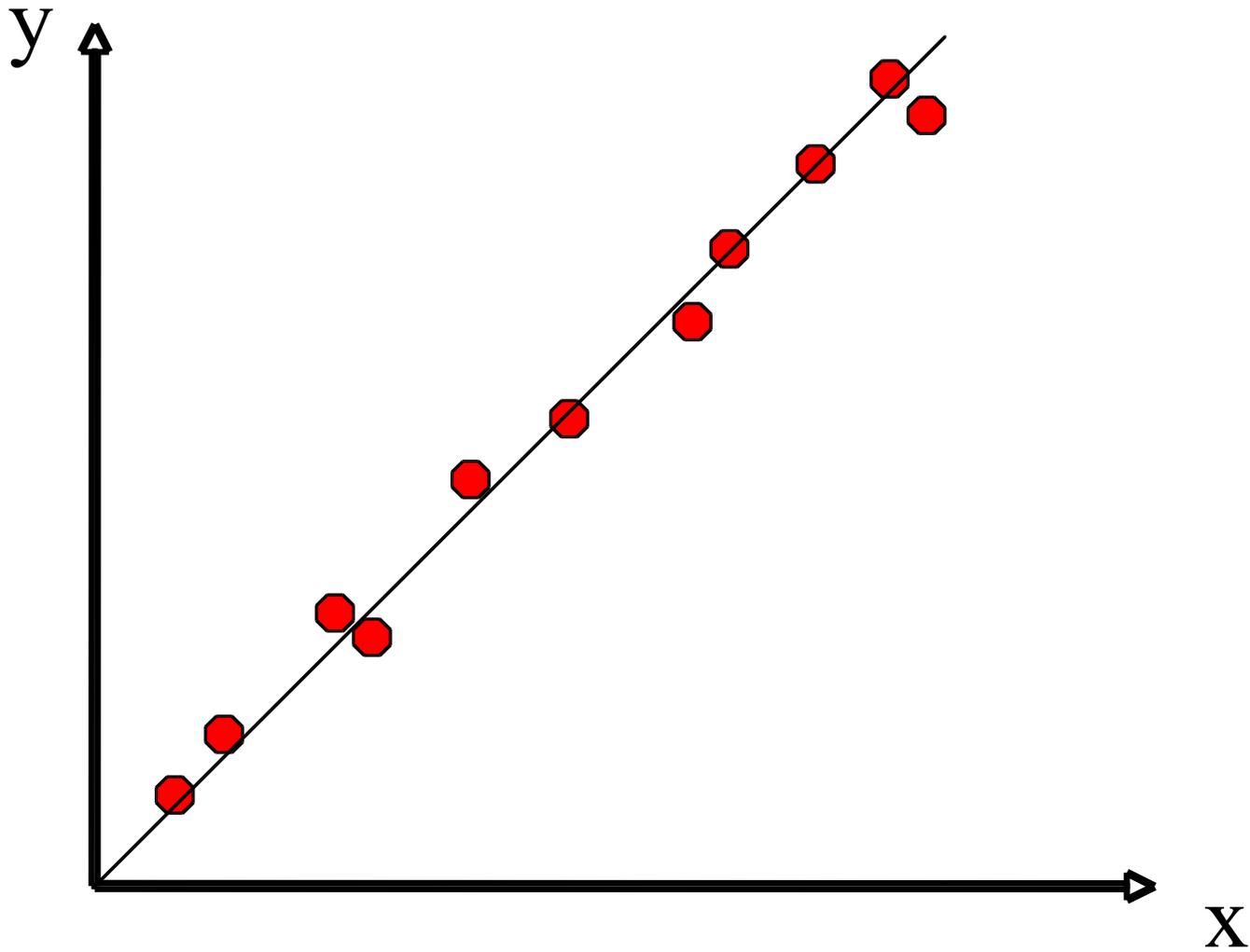


Esempio di output di un'analisi di classificazione "supervisionata": grafico di dispersione dei campioni di tre classi predeterminate (tre cultivar di vino) nello spazio definito dalla analisi discriminante lineare (**Linear Discriminant Analysis**) in base a Misure di etanolo, acido malico, ceneri, alcalinità, magnesio, fenoli, flavanoidi, non flavonoidi, proantocianine, colore, tonalità, prolina.

Metodi di Classificazione

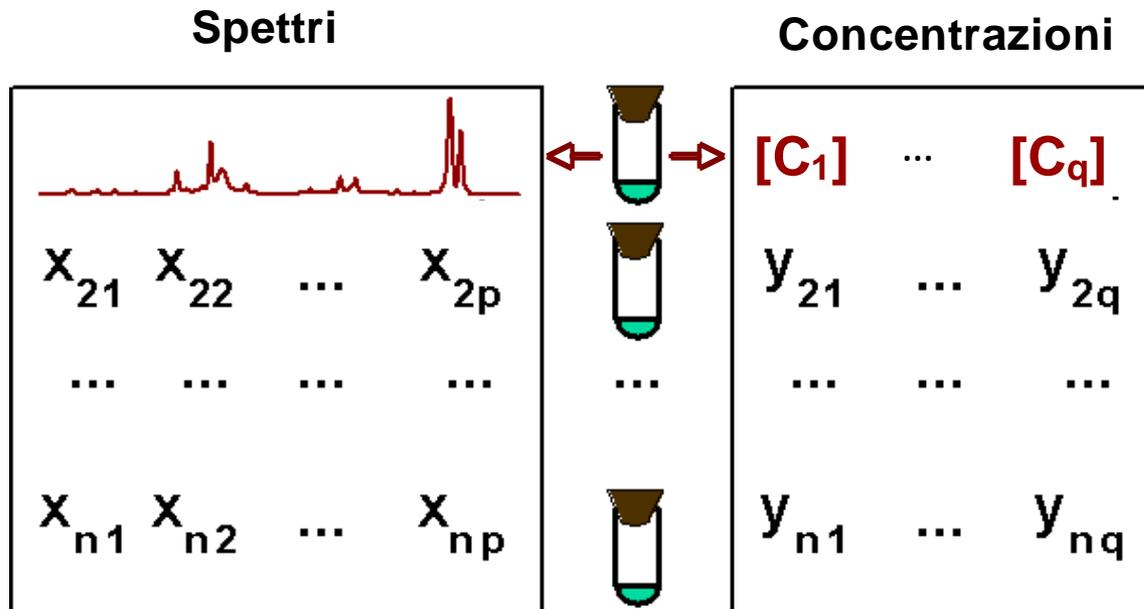
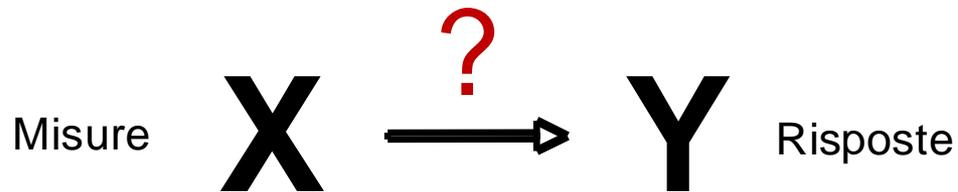
4) **Regressione / Calibrazione**

- Due tipi of variabili X / y
- Relazioni lineari / nonlineari
- Modelli
- Analisi diagnostica sulla bontà del modello



Regressione / Calibrazione⁷⁰

Calibrazione multivariata

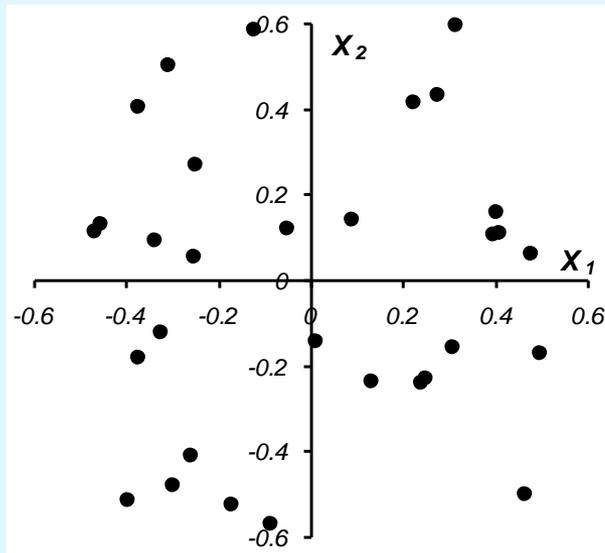


Un “*working horse*” per l’analisi esplorativa, la compressione dell’informazione e la visualizzazione di dati multivariati:

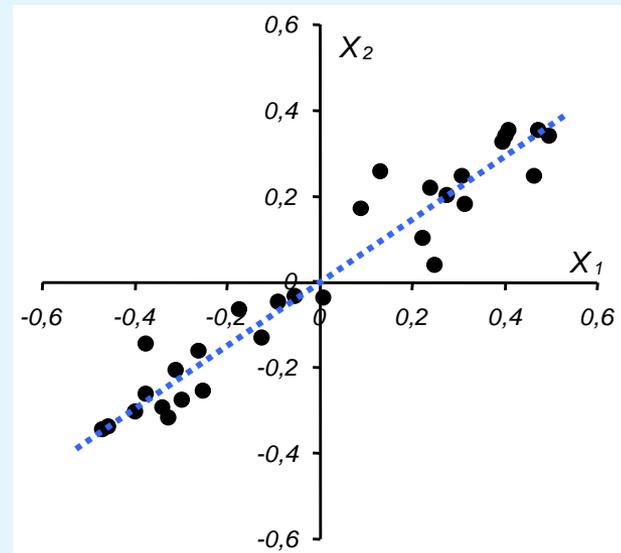
l’Analisi delle Componenti Principali (Principal Component Analysis - PCA)

Metodi di proiezione e variabili latenti

Dati senza struttura



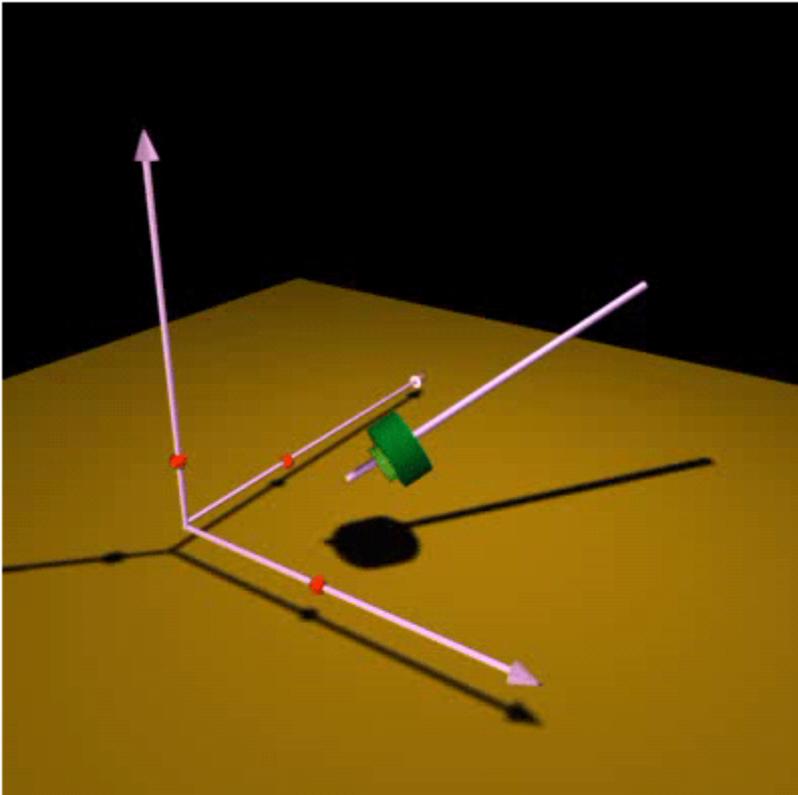
Dati con struttura nascosta



Metodi di proiezione e variabili latenti

Dimensioni formali – numero di variabili

Dimensioni effettive – numero di variabili latenti che coprono tutta la variabilità dei dati



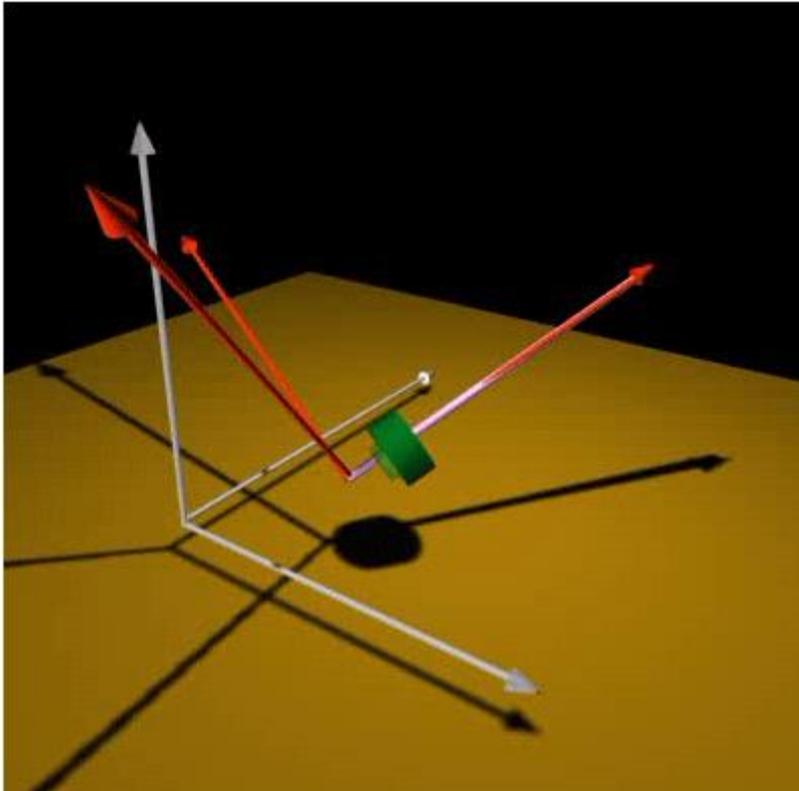
Dimensione formale = 3

X_1	X_2	X_3
0,121	0,095	0,259
0,834	0,951	0,901
1,548	1,807	1,543
2,261	2,663	2,185
2,974	3,519	2,827
3,687	4,375	3,469
4,401	5,231	4,111
5,114	6,087	4,753
5,827	6,943	5,394

Metodi di proiezione e variabili latenti

Dimensioni formali – numero di variabili

Dimensioni effettive – numero di variabili latenti che coprono tutta la variabilità dei dati



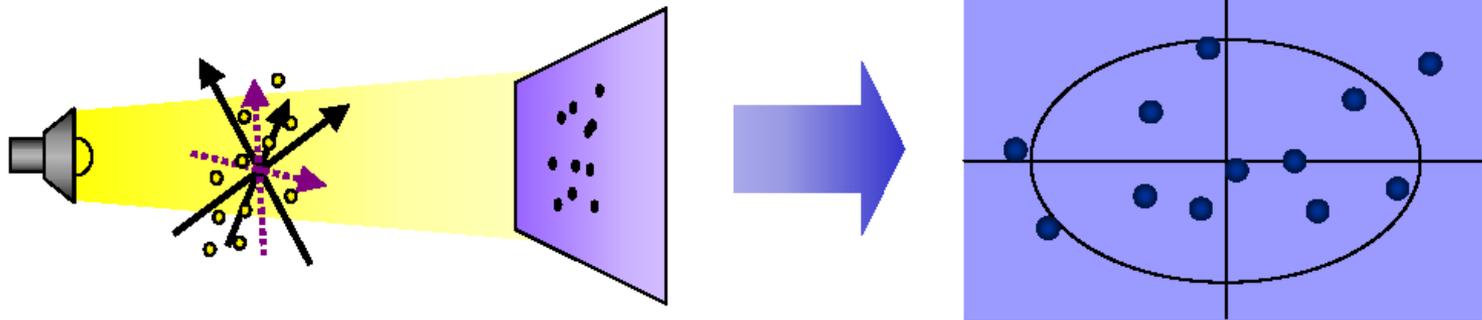
Dimensione effettiva = 1

X_1'	X_2'	X_3'
0,1	0,0	0,0
0,2	0,0	0,0
0,3	0,0	0,0
0,4	0,0	0,0
0,5	0,0	0,0
0,6	0,0	0,0
0,7	0,0	0,0
0,8	0,0	0,0
0,9	0,0	0,0

Metodi di proiezione e variabili latenti

Proiezioni nel sottospazio delle variabili latenti

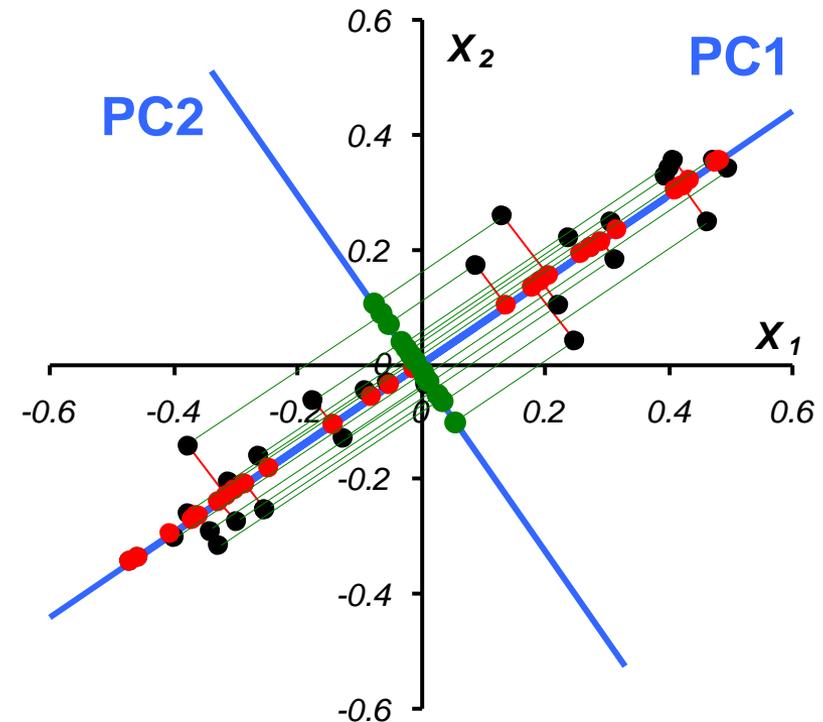
- Consente di ridurre la dimensionalità dei problemi
- Fornisce la possibilità di un'analisi visuale dei dati



Come trovare le variabili latenti?

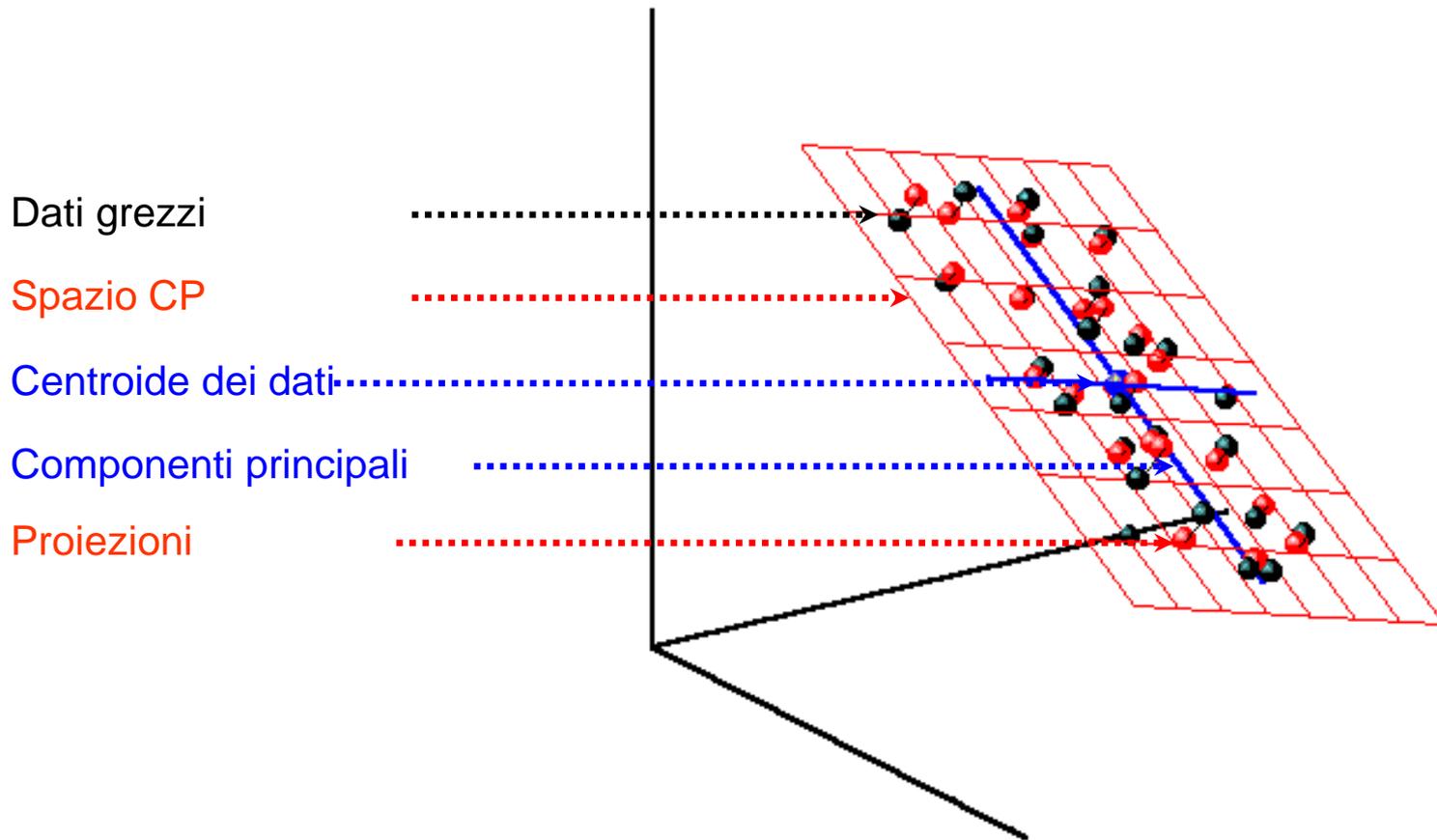
Spazio delle Componenti Principali

- Si individua la variabile latente – (prima **componente principale**, **PC1**) lungo la direzione di massima varianza
- Si proiettano tutti i campioni su **PC1**
- Rimane della **varianza residua**
 - considerata come noise/ rumore (informazione inutile)
 - modellabile con **PC2**

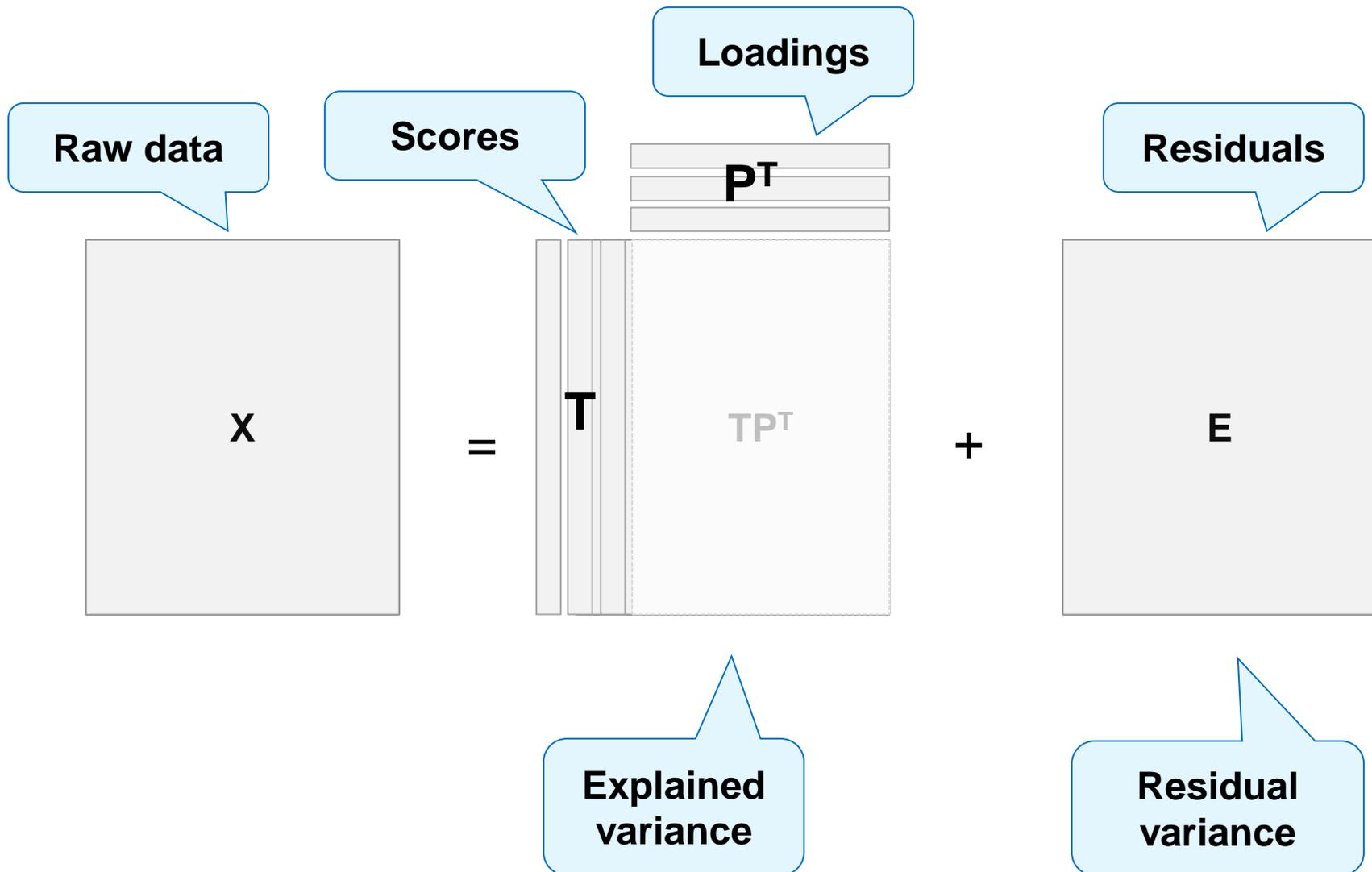


$$X_2 = aX_1 + E$$

Spazio delle Componenti Principali

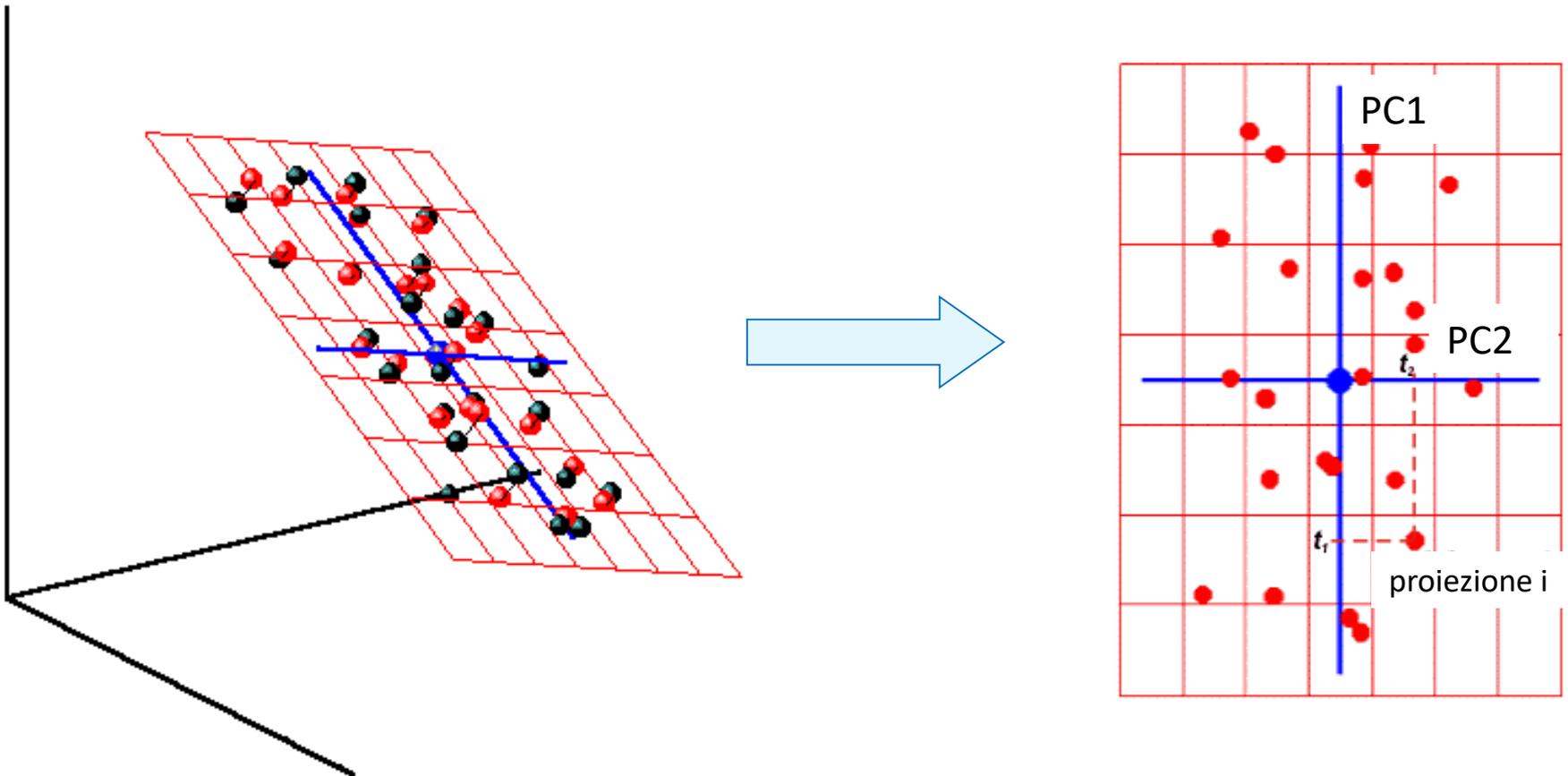


Analisi delle Componenti Principali



Punteggi (*scores*) nella PCA

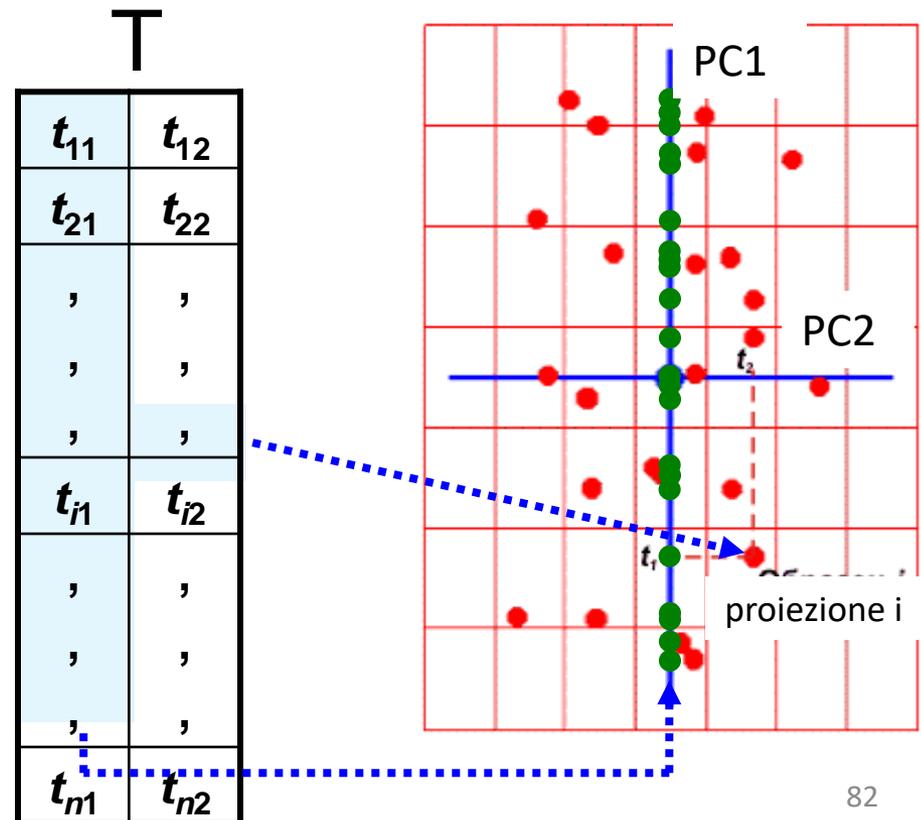
$$X = TP^T + E$$



Punteggi (scores) nella PCA

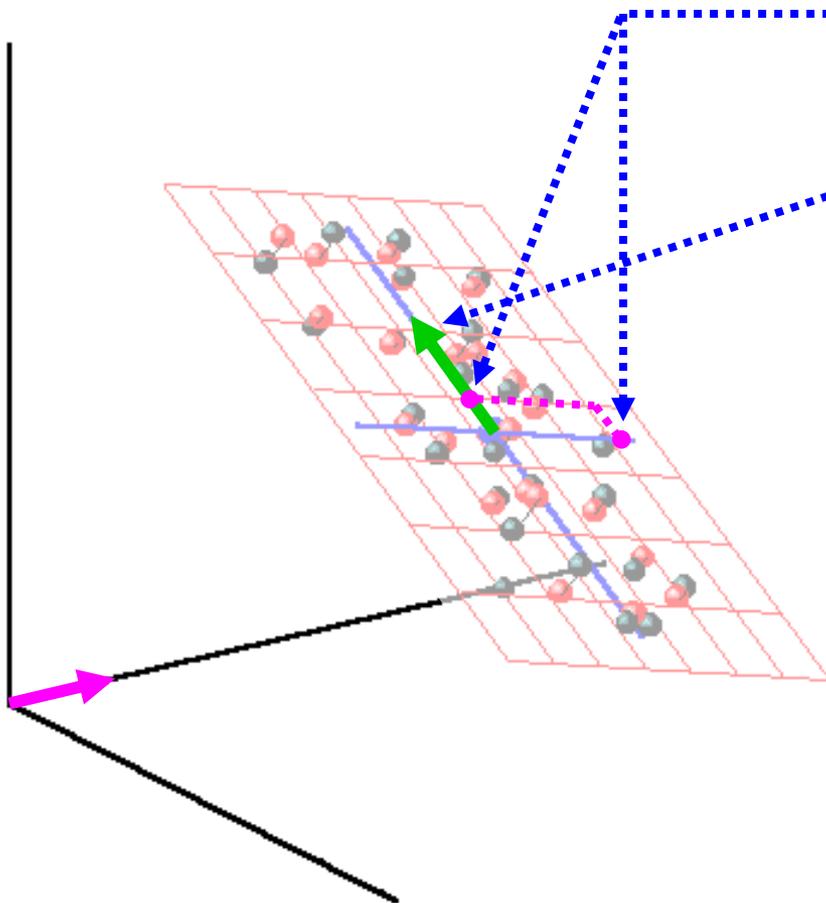
$$X = TP^T + E$$

- **Righe** – coordinate del campione sulle componenti principali
- **Colonne** – proiezioni dei campioni sulla componente principale



Pesi (*loadings*) nella PCA

$$X = TP^T + E$$



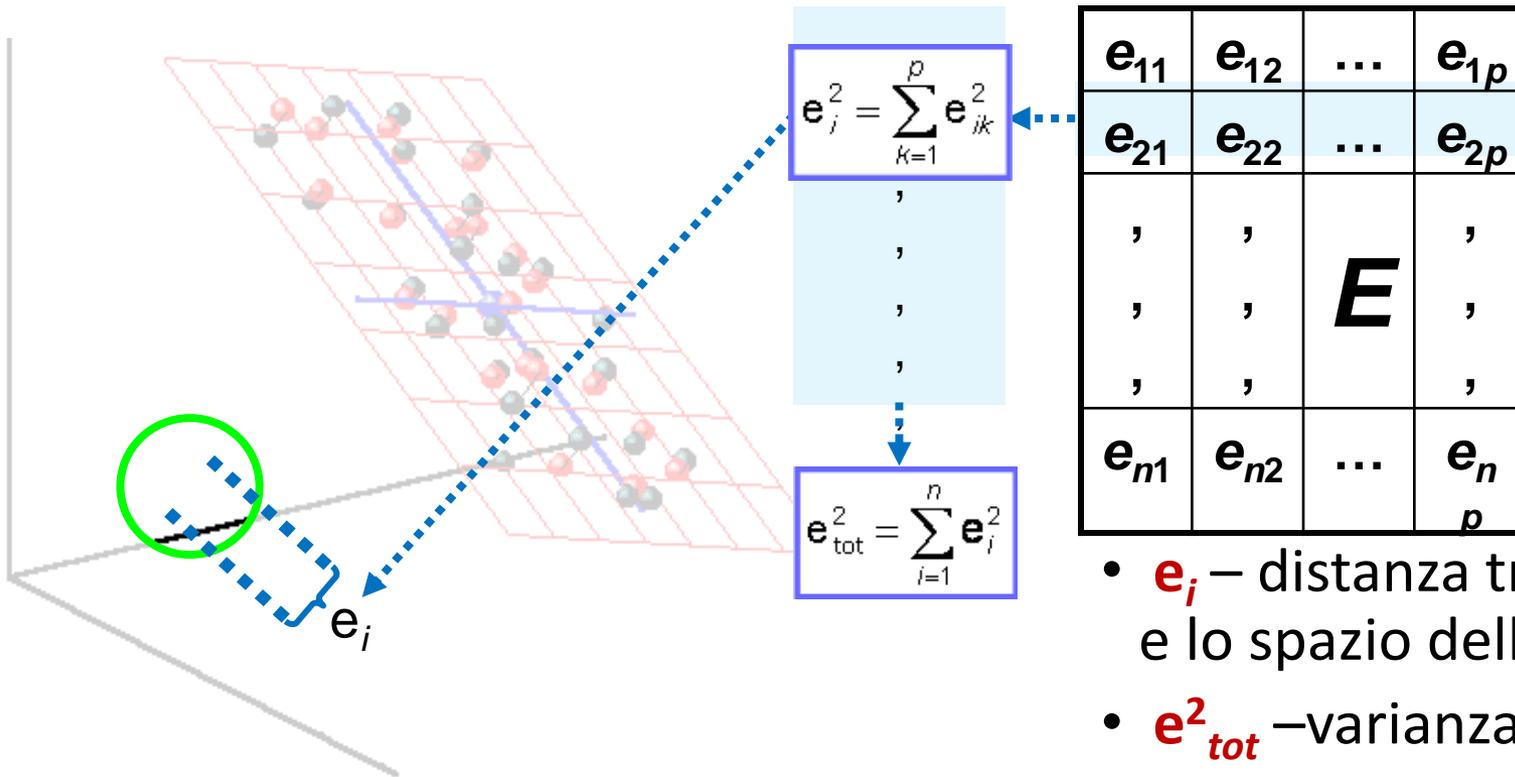
P^T

p_{11}	p_{11}	\dots	p_{1m}
p_{21}	p_{21}	\dots	p_{2m}

- **Righe** – vettori della base delle C.P. nello spazio delle variabili
- **Colonne** – proiezioni dei vettore di base delle variabili nello spazio delle C.P.

Matrice E, variabilità non spiegata dal modello

$$X = TP^T + E$$



- e_i – distanza tra il campione e lo spazio delle C.P.
- e_{tot}^2 – varianza residua

L'Analisi delle Componenti Principali
è una particolare tecnica di Analisi Fattoriale

Variabili e fattori

- ◆ L'analisi fattoriale ha come obiettivo principale l'individuazione di pochi costrutti fattoriali in grado di sostituire un insieme di numerose variabili.
- ◆ I costrutti fattoriali vengono considerati, a loro volta, nuove variabili suscettibili di una appropriata interpretazione

Caratteristiche comuni ai metodi di estrazione dei fattori

- ◆ Si estraggono fattori (comunque fino a un massimo pari al numero di variabili) finché si ritiene che la varianza spiegata sia sufficientemente grande rispetto alla varianza totale.
- ◆ Qualunque metodo di estrazione deve fornire, per ciascuna variabile, un valore numerico (chiamato "saturazione" o "peso") che misuri l'importanza del legame tra variabile e fattore.

Misura del grado di rappresentatività dei fattori rispetto a una variabile

- ◆ Per ciascuna variabile è possibile calcolare la "comunalità", cioè la somma dei quadrati dei "pesi" dei fattori.
- ◆ La "comunalità" assume valori compresi fra 0 e 1: se è uguale a 1, la variabile può essere esattamente determinata dalla combinazione lineare dei fattori.
- ◆ L'interpretazione dei risultati viene facilitata dalla "rotazione" della tabella dei "pesi", operazione per la quale sono disponibili diverse tecniche

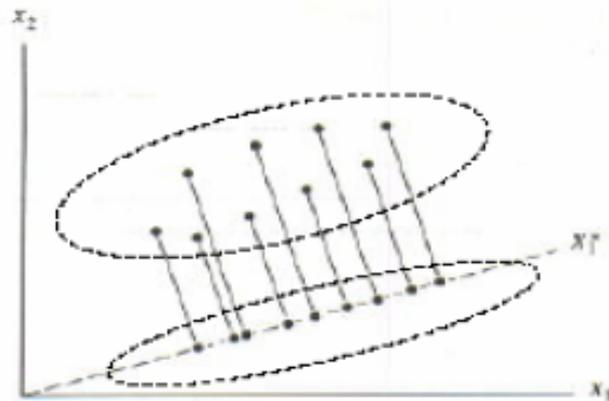
Analisi in componenti principali

L'Analisi in Componenti Principali (ACP) è una tecnica che a partire da un insieme di variabili quantitative (o al più binarie) osservate (originarie) $X_1, X_2, \dots, X_j, \dots, X_k$ produce un nuovo insieme di variabili artificiali Y_1, Y_2, \dots, Y_p ($p \leq k$) dove ciascuna Y_q ($q=1, \dots, p$) è una combinazione lineare di $X_1, X_2, \dots, X_j, \dots, X_k$

ACP da un punto di vista geometrico

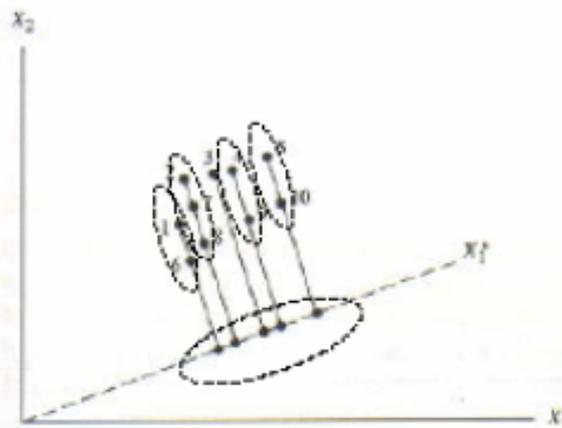
- La matrice dei dati $\mathbf{X}_{n,k}$ è rappresentabile geometricamente come n punti in uno spazio \mathbb{R}^k , cioè a k dimensioni
- Ciascuna riga della matrice $\mathbf{X}_{n,k}$ è chiamata vettore $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ik})$ costituito da k elementi numerici che rappresentano le coordinate cartesiane di un punto nello spazio di dimensione k
- L'ACP proietta gli n punti \mathbf{x}_i rappresentabili nello spazio \mathbb{R}^k in un sottospazio \mathbb{R}^p di dimensione ridotta $p < k$ individuato in modo tale che la "nuvola" degli n punti di \mathbb{R}^k sia deformata il meno possibile

Qualità della riduzione



Buona riduzione

Discriminazione tra le
unità preservata



Pessima riduzione

Formazione di cluster
indesiderati

Procedura

Obiettivo: si vuole individuare il sottospazio di dimensione 1, e cioè una retta \mathbb{R}^1 , tale che la proiezione degli n punti \mathbf{x}_i su di essa sia deformata il meno possibile

- Equivale a fare in modo che la varianza (l'inerzia) degli n punti \mathbf{x}_i proiettati sulla retta sia la più grande possibile
- Le coordinate degli n punti-proiezione y_{i1} sulla retta costituiscono i valori della variabile artificiale Y_1 costruita dall'ACP: questa variabile è chiamata *prima componente principale* e tali valori sono le "osservazioni" di tale componente principale

Calcolo della prima componente principale

■ Prima componente principale

$$y_{i1} = a_{11}x_{i1} + a_{21}x_{i2} + \dots + a_{k1}x_{ik} = \sum_{j=1}^k a_{j1}x_{ij} \quad i = 1, \dots, n$$

- Problema: determinare i valori dei coefficienti $a_{11}, a_{21}, \dots, a_{k1}$ in modo tale che la varianza della variabile artificiale (componente principale) Y_1 sia massima e che la somma dei quadrati dei coefficienti sia = 1

$$\text{Var}(Y_1) = \max$$

$$\sum_{j=1}^k a_{j1}^2 = 1$$

Calcolo delle componenti principali successive

■ Seconda componente principale

$$y_{i2} = a_{12} x_{i1} + a_{22} x_{i2} + \dots + a_{k2} x_{ik} = \sum_{j=1}^k a_{j2} x_{ij} \quad i = 1, \dots, n$$

- Problema: determinare i valori dei coefficienti a_{12} , a_{22}, \dots, a_{k2} in modo tale che la varianza della variabile artificiale (componente principale) Y_2 sia massima, che la somma dei quadrati dei coefficienti sia =1 e che la comp.princ. Y_2 sia incorrelata con la comp.princ. Y_1

$$\text{Var}(Y_2) = \max$$

$$\sum_{j=1}^k a_{j2}^2 = 1$$

$$r(Y_1, Y_2) = 0 \Rightarrow a_{11} a_{12} + a_{21} a_{22} + \dots + a_{k1} a_{k2} = 0$$

Proprietà delle componenti principali

- Ciascuna componente principale è una combinazione lineare delle variabili originarie
- La 1^a cp spiega il massimo della varianza (inerzia) spiegabile attraverso una riduzione ad una dimensione
- La 1^a componente e la 2^a componente principale spiegano il massimo della varianza spiegabile attraverso una riduzione a due dimensioni
-
- La 1^a componente, la 2^a componente, la p -esima componente principale spiegano il massimo della varianza spiegabile attraverso una riduzione a p dimensioni

$$\text{Var}(Y_1) \geq \text{Var}(Y_2) \geq \dots \geq \text{Var}(Y_p)$$

$$r(Y_t, Y_q) = 0 \quad \forall t, q \text{ tale che } t \neq q$$

Risultati dell'ACP

- La risoluzione del problema di massimo vincolato ad ogni passo porta a trovare che

✓ $\text{Var}(Y_1) = \lambda_1$ primo autovalore della matrice S di var. e covar.

Coefficienti $a_{11}, a_{21}, \dots, a_{k1}$ sono l'autovettore associato a λ_1

✓ $\text{Var}(Y_2) = \lambda_2$ secondo autovalore della matrice S

Coefficienti $a_{12}, a_{22}, \dots, a_{k2}$ sono l'autovettore associato a λ_2

✓ $\text{Var}(Y_3) = \lambda_3$ terzo autovalore della matrice S

Coefficienti $a_{13}, a_{23}, \dots, a_{k3}$ sono l'autovettore associato a λ_3

.....

✓ $\text{Var}(Y_p) = \lambda_p$ p -esimo autovalore della matrice S

Coefficienti $a_{1p}, a_{2p}, \dots, a_{kp}$ sono l'autovettore associato a λ_p

$$\text{Var}(Y_1) \geq \text{Var}(Y_2) \geq \dots \geq \text{Var}(Y_p) \Rightarrow \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$$

La matrice di correlazione R

$$\mathbf{R} = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1k} \\ r_{21} & 1 & \cdots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \cdots & 1 \end{pmatrix}$$

Le variabili vengono numerate da 1 a k, in modo che si potrà indicare con il simbolo r_{ij} il coefficiente di correlazione tra la variabile i e la variabile j . Data la simmetria del coefficiente di correlazione, si avrà $r_{ij} = r_{ji}$ ($i, j = 1, \dots, k$)

Risultati dell'ACP applicata a variabili standardizzate

- Se si vuole lavorare su variabili standardizzate allora la procedura descritta deve essere applicata alla matrice **R** di correlazione
- La risoluzione del problema di massimo vincolato applicato alla matrice **R** ad ogni passo porta a trovare che
 - ✓ $\text{Var}(Y_1) = \lambda_1$ primo autovalore della matrice **R** di correlazione
Coefficienti $a_{11}, a_{21}, \dots, a_{k1}$ sono l'autovettore associato a λ_1
 - ✓ $\text{Var}(Y_2) = \lambda_2$ secondo autovalore della matrice **R**
Coefficienti $a_{12}, a_{22}, \dots, a_{k2}$ sono l'autovettore associato a λ_2
.....
 - ✓ $\text{Var}(Y_p) = \lambda_p$ p -esimo autovalore della matrice **R**
Coefficienti $a_{1p}, a_{2p}, \dots, a_{kp}$ sono l'autovettore associato a λ_p

$$\text{Var}(Y_1) \geq \text{Var}(Y_2) \geq \dots \geq \text{Var}(Y_p) \Rightarrow \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$$

Le componenti principali sono fra loro incorrelate

Il modello ACP garantisce che le nuove variabili (le componenti principali) Y_1, Y_2, \dots, Y_p siano tra loro tutte non correlate

$$r(Y_t, Y_q) = 0 \quad \forall t, q \text{ tale che } t \neq q$$

Obiettivi dell'analisi in componenti principali

- Riduzione delle k variabili osservate in un numero inferiore $p < k$ di nuove variabili sintetiche dette componenti principali tra loro incorrelate tali che spieghino il massimo della varianza totale della nuvola di punti originaria
- Eliminazione della correlazione esistente tra le k variabili originarie osservate sostituendo ad esse le componenti principali che sono incorrelate
- Costruire indici sintetici e variabili sintetiche

Criteri per la scelta del numero di componenti principali

Quante componenti principali scegliere?

- La scelta deve essere fatta in base a
 - ✓ Criterio di parsimonia: numero minimo possibile di componenti principali
 - ✓ Minima perdita di informazione
 - ✓ Minima deformazione nella qualità della rappresentazione

Scelta del numero di componenti principali

■ Criteri di scelta

- ✓ Percentuale di varianza (inerzia) spiegata dalle componenti principali almeno superiore al 70%

$$\frac{\lambda_1 + \lambda_2 + \dots + \lambda_p}{\text{varianza totale}} = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_p}{\sum_{j=1}^k \lambda_j}$$

- ✓ Analisi della rappresentazione grafica degli autovalori delle componenti in ordine decrescente: si può disegnare una spezzata unendo i punti corrispondenti agli autovalori per individuare più facilmente le componenti davvero importanti
- ✓ Componenti principali corrispondenti ad autovalori λ_q il cui valore
 - ✓ è maggiore dell'"inerzia" media $(\lambda_1 + \lambda_2 + \dots + \lambda_k)/k$, OPPURE
 - ✓ è maggiore di 1 se nell'ACP si considerano le variabili standardizzate (analisi condotta sulla matrice di correlazione)

La matrice dei pesi (factor loading) delle componenti principali

- La q -esima componente principale Y_q è definita come la combinazione lineare delle variabili originarie $X_1, X_2, \dots, X_j, \dots, X_k$ con coefficienti $a_{1q}, a_{2q}, \dots, a_{jq}, \dots, a_{kq} \Rightarrow$ il generico coefficiente a_{jq} rappresenta il peso che la variabile X_j ha nella determinazione della c.p. Y_q
 - ✓ Valore assoluto di a_{jq} individua l'importanza della variabile X_j nella determinazione e spiegazione della cp Y_q
 - ✓ Il segno (positivo o negativo) di a_{jq} fornisce indicazione della relazione esistente tra X_j e Y_q

Calcolo della comunaltà di una variabile

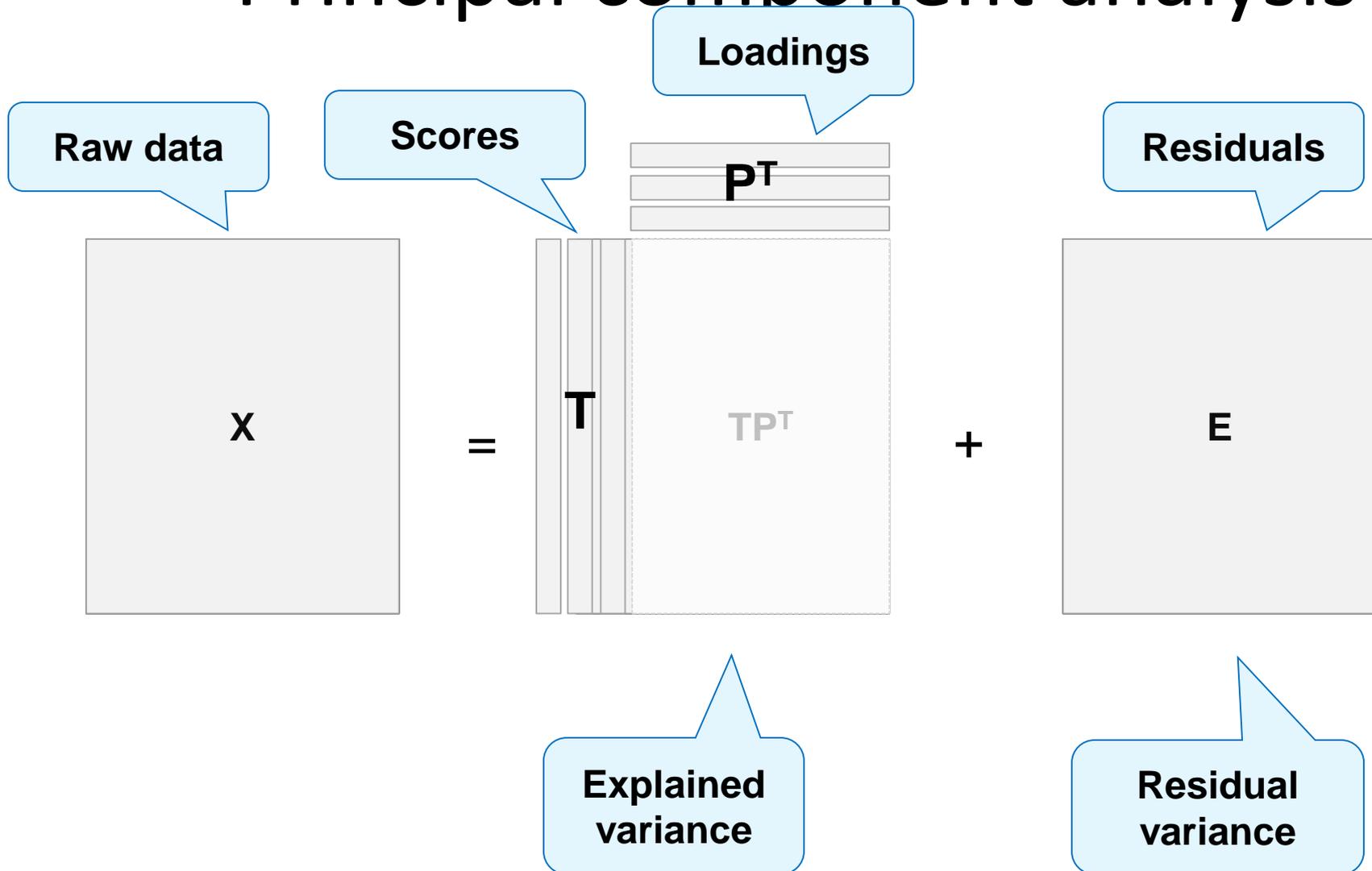
Assumendo p componenti principali, la comunaltà della variabile X_j è:

$$h_j^2 = (a_{j1} \sqrt{\lambda_1})^2 + (a_{j2} \sqrt{\lambda_2})^2 + \dots + (a_{jp} \sqrt{\lambda_p})^2$$

$$j = 1, 2, \dots, k$$

La comunaltà indica in quale misura le p componenti principali prescelte sono in grado di rappresentare ciascuna delle variabili originali.

Principal component analysis



Il processo analitico: generalità (capitolo 4 Analytical Chemistry)

- Il processo di misura chimico e suoi stadi: campionamento; preparazione del campione; misura e trasduzione del segnale analitico; acquisizione del segnale ed elaborazione dei dati.

Analisi elementare (capitolo 24 Analytical Chemistry)

- Spettrometria di emissione atomica; principi, sorgenti; spettrometri; rilevazione; prestazioni analitiche
- Spettrometria di assorbimento atomico; principi; sorgenti di radiazione primaria; sorgenti di atomi liberi; sistemi ottici dispersivi; rilevatori; misura dei segnali; sensibilità; interferenze chimiche; interferenze spettrali
- Spettrometria di fluorescenza di raggi X; principi, strumentazione; applicazioni
- Spettrometria di massa inorganica

Analisi specifiche su composti o molecole (capitolo 25 Analytical Chemistry)

- Spettroscopie molecolari UV-Vis e IR: principi; spettrofotometria UV-Vis, legge di Lambert-Beer; spettrofotometro; sorgenti; selettori di lunghezze d'onda; rivelatori, applicazioni in analisi qualitativa e quantitativa
- Spettrometria di massa organica: principi; un semplice spettrometro di massa; tecniche di ionizzazione soft; analisi delle masse; sistemi di introduzione del campione; strumentazione; spettrometria di massa in tandem; performance analitiche; analisi qualitative e quantitative; applicazioni

Sistemi eterogenei

Sistemi Liquido-Liquido (capitolo 14 Analytical Chemistry)

- Introduzione; Costanti di distribuzione: il coefficiente di partizione, rapporti di distribuzione; Estrazione di specie molecolari; estrazione di complessi metallici; Reagenti per l'accoppiamento ionico (ion pairing); Chelanti per i metalli

Equilibri Liquido-Solido (capitolo 15 Analytical Chemistry)

- Scambio ionico; Estrazione Soxhlet; Estrazione con fluidi supercritici; Accelerated solvent extraction (ASE); Metodi di estrazione assistiti dalle microonde (con solventi organici e digestioni acide); Estrazioni ultrasoniche; Estrazioni con acqua supercritica // Estrazione con fase solida (SPE); Metodi di estrazione per assorbimenti ("sorbitive") (SPME, SBSE)

Sistemi Gas-Liquido e Gas Solido (capitolo 16 Analytical Chemistry)

- Sistemi Gas-Liquido; Sistemi Gas Solido

PROGRAMMA 2020-2021 (3)

Cromatografia (capitolo 21 Analytical Chemistry)

- Fondamenti delle separazioni cromatografiche; sviluppo di un cromatogramma; valori caratteristici di un cromatogramma; teoria della cromatografia; La risoluzione R_s come parametro della separazione dei picchi; analisi qualitativa e quantitativa.
- La gas-cromatografia; dati sulla ritenzione e coefficienti di partizione; separazioni nella fase gassosa; componenti di un gas cromatografo; fasi stazionarie; applicazioni; cromatografia di adsorbimento.
- La cromatografia liquida; high performance liquid chromatography (HPLC); fasi legate; cromatografia di adsorbimento; cromatografia ionica classica e HPIC; esclusione dimensionale; cromatografia su strato sottile;
- Cromatografia con fluidi supercritici; tecniche multidimensionali; elettroforesi, *Field Flow Fractionation*.

Accoppiamento di tecniche cromatografiche e spettroscopiche (capitolo 26 Analytical Chemistry)

- Introduzione; sistemi gas cromatografici ibridati; GC-MS; LC-MS; altre tecniche ibridate

Metodi elettrochimici - Sensori chimici (capitoli 18 e 33 Analytical Chemistry)

Preparazione del campione (capitolo 20 Analytical Chemistry)

- Introduzione; macinazione, omogeneizzazione ed essiccazione del campione; dissoluzione e digestione di specie insolubili; filtrazione e tecniche di pretrattamento del campione basate su membrane; tecniche di spazio di testa; estrazione; tecniche di estrazione liquida; intrappolamento su un solido (ad)sorbente; estrazione di analiti inorganici; procedure cromatografiche come separazioni preliminari.

Il processo analitico: ricapitolazione ed approfondimento (capitolo 4 Analytical Chemistry)

- Il processo di misura chimico; Operazioni preliminari; Il campionamento (fattori limitanti, approcci, tipi di campione, errori, metodi); La preparazione del campione; La misura e la trasduzione del segnale analitico; L'acquisizione del segnale e l'elaborazione dei dati; *Introduzione ai metodi chemiometrici*. Tendenze attuali scientifiche e tecniche; esempi di processi analitici

Previsti 3 CFU «Ottimizzazione sperimentale, analisi e visualizzazione di dati chimici, fisici e biologici» con finanziamento regionale (Prof.S. Licen ? pausa tra gennaio-marzo 2021)

ESAME FINALE

Esame finale:

- Esame orale integrato (teoria e laboratorio);
- Discussione delle relazioni sulle esperienze di laboratorio in sede di esame.

Relazioni su esperienze di laboratorio:

- Le relazioni su tutte le esperienze di laboratorio verranno consegnate dagli studenti ai docenti entro 7-10 giorni prima della data dell'esame;

- I docenti renderanno disponibile su piattaforma Moodle2 il materiale (*slides*) riguardante gli argomenti trattati durante le lezioni;